

# THIS WEEK



## EDITORIALS

**NEW DEAL** Welcome the growing band of venture philanthropists **p.266**

**WORLD VIEW** China must improve the quality of its research **p.267**

**GLORIOUS** Technicolour toad resurfaces in Borneo jungle **p.270**

## Don't blame the CIA

*Tales of fake vaccination drives are the least of Pakistan's public-health problems. A disjointed care system and lack of services are doing more damage.*

The story is so far-fetched that it might just be true. According to newspaper reports, US spies this year conscripted a Pakistani doctor to run a fake vaccination campaign in the quiet city of Abbottabad, as part of a plot to uncover the hiding place of Osama Bin Laden. The Central Intelligence Agency (CIA) supposedly wanted to confirm the whereabouts of the United States' most-wanted terrorist before he was targeted by US special forces in a raid on the town in May (see *Nature* <http://dx.doi.org/10.1038/news.2011.418>; 2011).

Within days of the story's publication in UK newspaper *The Guardian*, the CIA was attacked by Pakistani doctors, public-health experts working in the country and Médecins Sans Frontières, a major aid agency based in Geneva, Switzerland. They warned that stories of the sham vaccination drive, which targeted hepatitis B, would reinforce conspiracy theories already circulating in Pakistan, and could drive up refusal rates for all vaccines — just when the country's efforts to eradicate polio through vaccinations are flagging.

The CIA has refused to confirm or deny the story, but the reaction to its supposed plot is telling. Pakistan's investment in health care is among the lowest in the world — just 2.6% of gross domestic product. Infant-mortality rates are high, and health-service coverage is poor. Against this desperate backdrop, few familiar with the nation's health system would doubt suggestions that a government doctor could be bought off by a foreign intelligence service to aid in an assassination plot. Nor are they likely to be surprised at the same doctor then having travelled to a major city, paid off low-level health-care workers and conducted a mass vaccination campaign without detection.

The scheme has hit a nerve because, until now, vaccinations have been among the few bright spots in Pakistan's health statistics. Too many children still go unprotected, but outright refusals of polio vaccine by parents account for just 6% of missed children, according to the United Nations Children's Fund. This is despite the anti-vaccine conspiracy theories that had been circulating long before stories of the CIA plot emerged. In recent years, vaccination rates for other diseases, such as measles, have been around 80–85%, not far behind developed nations such as the United Kingdom.

That is not to understate the challenges. Polio virus is surging in Pakistan, and 2010 saw a 62% increase in cases over the previous year. This year, the World Health Organization has already reported 59 cases of children paralysed by the virus — almost double the number at this time last year. Many cases are in the Federally Administered Tribal Areas, a semi-autonomous region that is difficult to reach, in part because of ongoing military activity. But the southern city of Karachi has also seen an increase in polio cases, attributable in part to unvaccinated refugees arriving in the city after being displaced by last summer's flooding.

A report last month by the Independent Monitoring Board of the Global Polio Eradication Initiative in Geneva noted that Pakistan's broken health system is a major factor in the disease's resurgence. Leadership from the top is weak, and under this year's devolution of

the system to local governments, it looks set to become even weaker.

Meanwhile, reports from researchers on the ground suggest that the morale of the nation's army of 'volunteer' vaccinators, paid less than US\$2 a day, is flagging. The volunteers say that the people they are trying to help are increasingly hostile to their efforts, having been subjected to a decade of vaccination drives but offered few other health services.

The situation is similar to that in northern Nigeria, where vaccination rates have plummeted in the past decade because of strife and

**"Volunteers say that the people they are trying to help are increasingly hostile to their efforts."**

armed conflict. Scare stories that vaccinations are part of a Western plot to sterilize Muslims posed (and continue to pose) significant obstacles to polio eradication. Trust is being rebuilt in Nigeria, in part by charities and the national government working with local leaders, but also through improvements in overall health care. For example,

as incentives to allow vaccination, families were offered bed nets to deter mosquitoes carrying the much-feared disease malaria. Polio rates have been reduced by 95% over the past year, although much work remains.

The latest reports of a CIA-led vaccination plot are troubling and could complicate public-health campaigns in Pakistan at a key time. But they will not be a deciding factor in the war against polio and other preventable diseases. Pakistan's decrepit and failing health-care system poses a far greater threat. And it deserves as much attention. ■

## Growing pains

*It is time to update decades-old regulation of genetically engineered crops.*

Researchers at Scotts Miracle-Gro have a vision of a greener future. The lawn-care company, based in Marysville, Ohio, wants to develop a dwarf grass that needs less frequent maintenance than standard Kentucky bluegrass. But there is a catch: such grass is unlikely to stand up to weeds. No problem, the company reasons, it will make a dwarf grass that is resistant to herbicide to help homeowners to nip those weeds in the bud.

Development of this genetically modified (GM) Kentucky bluegrass made headlines this month when the US Department of Agriculture (USDA) told Scotts that it did not have the authority to regulate it (see page 274). As a result, Scotts is free to start selling its new crop without oversight.

The reason for this is historical. US regulation of GM crops relies

on its authority to control plant pests, and so the USDA has regulated crops on the basis of the way plant-pest-based tools are used to make them. It is a bizarre approach, given the low pest risk from the tools. But it had some merit when it was first developed because foreign genes were often inserted into the plant genome by a bacterium that can be lethal to some plants. Once in place, the expression of the foreign gene was guided by a series of genetic elements pulled from plant viruses.

To get around this, researchers at Scotts made GM grass without using plant pests. It took more work, but the company reasoned that the streamlined regulation — as well as possible greater consumer acceptance and relief from the patent stranglehold on more traditional genetic-engineering methods — would make it worthwhile. So they mined the wealth of plant genomic data now available, snipped a herbicide-resistance gene from the model plant *Arabidopsis thaliana*, sewed it to genetic elements pulled from maize (corn) and rice to drive the gene's expression, and used a gene gun to blast it into the Kentucky bluegrass genome.

This technique is not the only GM method likely to fall outside USDA regulations. Plant biologists have made tremendous strides since the current rules were cobbled together in 1986, advancing both our fundamental understanding of plant genetics and the technical know-how in manipulating gene expression. Genetic changes can now be made at specific sites in the genome, and foreign genes can even be expressed in plant cells without integrating them into the genome at all. And gene expression can be regulated using RNA molecules — including, in some cases, ones made by the plant in response to attack by a pathogen.

Many of these advances are still years from commercialization. But regulators must prepare the ground. Monsanto GM soya beans, which use RNA interference to modulate the expression of endogenous genes, are already awaiting a decision from the USDA.

The USDA and others need to reconsider how they define and

control GM species. If a crop developer uses genetic engineering to delete a discrete segment of a plant genome, how much regulation does that require? Would those same guidelines be appropriate for a crop that expresses half-a-dozen foreign herbicide- and insect-resistance genes, engineered without the use of plant pests? Such questions are particularly important where — as in the United States — GM regulation rests not on the final product of genetic engineering, but on the methods used in the process.

The European Commission is tackling the issue, and has commissioned a study into how new plant techniques fall under the rubric of

***"In the United States, genetic-modification regulation rests not on the final product but on the methods used."***

the European Union definition of GM crops. Similarly, the USDA's Advisory Committee on Biotechnology and 21st Century Agriculture has raised the problem as a point of concern. But the USDA's proposed changes to its GM regulatory powers, released in draft form in 2008, failed to address challenges posed by new technologies.

The USDA's Kentucky bluegrass ruling comes at a crucial time for agricultural biotechnology. Some estimate that the world must increase the rate of growth in agricultural productivity by 25% per year to meet growing worldwide demand for food and biofuels. Many argue that advances in agricultural biotechnology, some of which may come from GM crops, will be needed to meet this demand. Industry, particularly smaller companies, needs to know how these crops will be regulated before they will invest to develop new techniques.

The new breed of GM crops could help gain wider acceptance for the technology, by settling long-standing unease about the use of foreign genes and the inability to target such genes to a specific location in the genome. But it is doubtful that dubious consumers are ready for GM crops to escape regulation altogether. ■

## With strings

*Researchers should shrug off their fears and welcome the concept of venture philanthropy.*

When the Maryland-based Cystic Fibrosis Foundation invested in Californian biotechnology company Aurora Biosciences in 2000, it launched a revolution. Before then, it was taboo for a biomedical charity to take a stake in a commercial firm; instead, foundations usually sent their money to academic labs. Those days are over — now is the era of 'venture philanthropy'.

Under this model, continued investment in research can depend on projects reaching predetermined milestones and deadlines. And, as we report on page 275, charities have started to take an interest in controlling the intellectual property that results from such projects. That idea makes some uneasy, but the benefits extend beyond royalties: clauses in intellectual-property agreements can be used to protect a philanthropic investment as well. One risk of working with industry, for example, is that a promising drug can be shelved if the company that owns the patent rights pulls the plug on efforts to develop it as a therapy. To protect against this, much research funded by philanthropies is now subject to interruption licences, which allow charities to regain — and relicence — intellectual-property rights if a project ceases.

Then there is the 'research-only' clause, which promotes continued scientific progress in a field by encouraging companies to allow academic labs to study patented technology. However, patents remain an important currency in business. The best way to develop a new drug is probably for charitable investors to take a guiding, but not overly controlling, hand in intellectual property. If a charity demands high

royalties, industrial partners — beholden to the financial demands of their investors — may shy away from the project. And if a charity calls for co-ownership of the intellectual property with a company or university, potential partners might hesitate to license the resulting patents.

University researchers can also benefit from paying closer attention to intellectual property. In some ways, the concept runs counter to the intellectual freedom prized in academia. But the venture-philanthropy approach could be a useful model, especially as the search for funds and the push towards translational research nudge more academic labs into partnerships with industry. Collaborations between academia and pharmaceutical companies are already using research agreements that grant researchers rights similar to those in interruption licences (see *Nature* 474, 433–434; 2011).

Many academics reject the notion of patents altogether, preferring their research to remain openly accessible. In some cases, this approach has worked. The Alzheimer's Disease Neuroimaging Initiative, a US-based public-private partnership, has unquestionably accelerated the search for new diagnostic tools without patenting its results. The Michael J. Fox Foundation has also taken this approach in its Progression Markers Initiative to find biomarkers of Parkinson's disease.

Industry has seen the value of such projects, and is pushing for more of them. But the approach works best when laying important, early-stage scientific groundwork. Eschewing patents can stifle the development of downstream projects by discouraging private-sector investment.

Yet that does not mean that academics — or charities — should capitulate completely to industry's demands. Indeed, both should expect some push-back from industry at the negotiating table on even

minor control measures such as interruption licences. But to take a stronger line on the ownership of intellectual property will ultimately help all those involved in health-care research to turn ideas into therapies. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunjv](http://go.nature.com/xhunjv)





## Focus on quality, not just quantity

*China publishes huge amounts of scientific research. Now it must make more of it worth reading, says Changhui Peng.*

China's recent rise to scientific superpower has been striking. A report published earlier this year by London's Royal Society found that China now publishes the second highest number of scientific papers and that, by 2020, it could be the world's dominant producer of scientific research.

China has intensified its investment in research and development in recent years. Spending has grown by 20% annually since 1999, and has now reached more than US\$100 billion a year. The Chinese government has urged scientists to publish in highly respected English-language journals, offering promotions and other rewards as incentives; and many Chinese universities have attempted to boost their rankings in the Shanghai Jiao Tong University's world university table, which is weighted heavily towards articles published in *Science* and *Nature*.

However, despite the enormous progress made in China during the past few decades, the quality of its research seems not to have kept pace. The Royal Society report used the number of times a paper is cited in the scientific literature as a proxy for quality. It found that between 1999 and 2008, China's citation share rose from almost nothing to 4%. However, this is dwarfed by the 30% share held by the United States. And although China ranks second to the United States in terms of publication output, the report found that, in 2008, it ranked only joint ninth in citation numbers. This suggests that China's dramatic proliferation of scientific papers does not reflect quality research. China still has a long way to go to become a major player in the scientific arena and, to do so, I believe it must address these key areas.

First, data sharing. Wide distribution of information is key to scientific progress, yet traditionally, Chinese scientists have not systematically released data or research findings, even after publication. With so much emphasis on publication, data sharing is regarded as less important, and rules to encourage or compel such behaviour are inadequate. Moreover, institutions want to monopolize their data in the interest of their future scientific reports.

There have been widespread complaints from scientists inside and outside China about this lack of transparency. What data are made routinely available are often satellite measurements made for meteorology or large-scale background Earth-systems science records. Usually incomplete and unsystematic, these data are of little value to researchers and there is evidence that this drives down a paper's citation numbers.

Alongside better data access, China must do more to monitor and punish widespread academic misconduct, including plagiarism, which occurs as a consequence of the emphasis placed on publishing large numbers of papers. The CrossRef service, offered by the nonprofit association

CrossRef, could help Chinese publishers to identify plagiarism, by comparing the content of a submitted paper to a continuously updated database of published work.

The third area that needs improvement is international collaboration. Fuelled by a desire to work with the best people, as well as by advances in communication technologies and more affordable travel, international scientific collaborations are on the rise. According to the Royal Society report, the past 15 years has seen a 10% increase in the number of published articles that are internationally collaborative. There is also a strong correlation between citation number and the number of collaborating countries (up to a tipping point of ten countries).

There is already progress here, and China is beginning to open up. The Chinese Ministry of Science and Technology has signed treaties

for scientific and technological cooperation with more than 100 countries. Under these treaties, the Chinese government is encouraging scientists to cooperate and exchange data with international organizations. China is also welcoming international scientists to come in and set up long-term cooperative initiatives. These efforts should be accelerated and their profile raised. Only by participating in more international scientific collaborations, such as the Intergovernmental Panel on Climate Change or the FLUXNET global network of micrometeorological tower sites, can China catch up with the United States and Europe.

The final area is the way in which China addresses complex and interrelated global issues — including climate change, Earth-systems modelling, carbon-capture technologies, biodiversity and resource security. To be a scientific superpower, China must encourage its scientists to play

a more prominent part in addressing these pressing challenges. Chinese scientists should think globally and put themselves at the forefront of cutting-edge science. They must demonstrate leadership, developing new research initiatives and chairing international programmes. A good example is the Third Pole Environment programme, led by the Chinese Academy of Sciences' Institute of Tibetan Plateau Research in Beijing, which aims to pool international resources and expertise to study the interactions between ice, water, air, ecology and human behaviour.

The time has come for China to consider how best to boost the quality, rather than the quantity, of its scientific output. The steps I have outlined will provide a platform to strengthen the impact of China's research and contribute valuable science to the world's most important questions. ■

ALTHOUGH CHINA  
RANKS SECOND  
IN TERMS OF  
**PUBLICATION**  
OUTPUT, IT RANKS  
ONLY NINTH IN  
**CITATION**  
NUMBERS.

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/ngfgy1](http://go.nature.com/ngfgy1)

**Changhui Peng** is at the College of Forestry of Northwest A&F University, Yangling, China, and the Institute of Environmental Sciences of the University of Quebec at Montreal in Canada.  
e-mail: [peng.changhui@uqam.ca](mailto:peng.changhui@uqam.ca)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## MICROBIOLOGY

### Bacteria 'blink' to expel molecules

A voltage-sensitive fluorescent protein has revealed, at the single-cell level, the electrical signals that bacteria use to eject compounds.

The electrical potential across biological membranes drives the transport of some molecules into and out of cells, but measuring this voltage difference in bacteria has proved difficult. Adam Cohen and his colleagues at Harvard University in Cambridge, Massachusetts, modified the marine bacterial protein proteorhodopsin so that it fluoresced in response to voltage changes. They then expressed the engineered protein in the bacterium *Escherichia coli*.

When the bacteria were exposed to a membrane-permeable dye, flashes of fluorescence coincided with precipitous decreases in the amount of dye in the cell. This suggests that the dye is pumped out of the cell in response to electrical signals. *Science* 333, 345–348 (2011)

## ANIMAL BEHAVIOUR

### Learning lizards make smart moves

Lizards have surprised researchers by demonstrating flexible problem-solving and learning skills previously seen

mostly in birds and mammals. The reptiles had been thought to have rigid, stereotyped behaviour patterns and limited cognitive abilities.

Manuel Leal and Brian Powell at Duke University in Durham, North Carolina, presented six Puerto Rican *Anolis evermanni* lizards with two wells (pictured), one of which contained a fly larva reward and was associated with a plain blue disc. After a habituation period, the creatures were challenged to dislodge the blue disc covering the well with the reward. Four of the six lizards repeatedly solved this problem by either biting or shoving the cap aside to reveal the treat, and chose

the blue disc over differently coloured discs. When the reward was placed under a new disc colour, two lizards were able to reverse their choice.

Such behavioural flexibility may have enabled *Anolis* lizards to radiate across the tropics of the Americas, and suggests that scientists should rethink their ideas on reptile cognition. *Biol. Lett.* doi:10.1098/rsbl.2011.0480 (2011)

## HIV

### Antibody search hits gold

A treasure trove of 576 antibodies that bind to and neutralize HIV has been

discovered in four infected individuals, vastly expanding the number of antibodies known to inactivate a broad range of HIV strains. Such molecules could be useful in treating, or even preventing, HIV infection.

Only a handful of broadly neutralizing antibodies against HIV had previously been isolated, partly because the molecules mutate so often. So Michel Nussenzweig at the Rockefeller University in New York and his colleagues devised a new way to fish out the antibodies — by targeting an area of the molecules not prone to frequent mutation.

They found that the



## ECOLOGY

### Carbon parks in the city

Urban green spaces lock up tonnes more carbon than previously thought.

By using satellite imagery and analysing vegetation carbon content, researchers estimate that around 230,000 tonnes of carbon are stored in the above-ground vegetation of Leicester (pictured) — an average-sized city in central England. This is equivalent to 3.16 kilograms of carbon per square metre of the city, an order of magnitude greater than current national

estimates for Leicester. The highest carbon-storage density was linked to tree cover in publicly owned or managed areas.

The team, led by Zoe Davies at the University of Kent in Canterbury, UK, recommends improved monitoring and management of urban vegetation to maximize its contribution to mitigating greenhouse-gas emissions.

*J. Appl. Ecol.* doi:10.1111/j.1365-2664.2011.02021.x (2011)





antibodies bind to gp120, an HIV surface protein that the virus uses to enter host cells. A subset of the antibodies neutralized 96% of 118 viruses in a test panel. Although the antibodies were highly mutated, they all shared a sequence of 68 amino acids.

**Science** doi:10.1126/science.1207227 (2011)

## VACCINES

## Mix-and-match for meningitis

A vaccine for meningitis has long eluded researchers because the key antigen of the meningococcus B strain has more than 300 sequence variations. Using the three-dimensional crystal structure of this antigen, fHBP, researchers have engineered an antigen that carries the main amino-acid variants and elicits antibodies against all strains of the bacterium in mice.

Rino Rappuoli at Novartis Vaccines and Diagnostics in Siena, Italy, Lucia Banci at the University of Florence, Italy, and their co-workers analysed the sequences of the 300 different types of fHBP. They classified the variants into three main groups and then engineered variants from one group to carry sequences from the other two. They tested 54 engineered fHBPs in mice and one stood out for its ability to induce the production of bacterium-killing antibodies. The authors say it could be used in a vaccine and that the approach could aid in the design of vaccines for other pathogens with many natural variants.

**Sci. Transl. Med.** 3, 91ra62 (2011)

## CHEMISTRY

## DNA-inspired polymerization

Polymerization typically relies on harmful metal catalysts, but researchers in Japan have succeeded in circumventing this problem. Akira Harada at Osaka University and his

colleagues constructed a synthetic polymerase that can catalyse the synthesis of high-molecular-weight polymers.

The polymerase is made up of two ring-shaped sugar molecules called cyclodextrins (CDs) linked together by a flexible covalent chain. The authors propose that one of the CDs functions as the active site, where key bonds in an incoming cyclic ester monomer are broken to open up the ester's ring, allowing it to bond with other monomers and form a chain. The second CD functions as a clamp, threading the growing chain through its hollow structure to hold the chain in place. As the polymerization proceeds, the growing chain slides by one position, freeing up space for the next incoming monomer.

The structure's design was inspired by the polymerases that synthesize DNA.

**Angew. Chem. Int. Edn** doi:10.1002/anie.201102834 (2011)

## CANCER

## Tissues stretch to let tumours move

A protein made by connective-tissue cells causes mechanical changes in tissue structure that help cancers to spread around the body.

Jacky Goetz and Miguel Del Pozo at the Spanish National Center for Cardiovascular Research in Madrid and their colleagues found that stromal fibroblast cells surrounding many human cancers express high levels of the protein CAV1. Mouse fibroblasts expressing CAV1 activate the enzyme Rho, which causes the cells to stretch out (pictured). In three-dimensional gel matrices *in vitro*, the elongated fibroblasts formed stiff, parallel-fibre networks through which cancer cells moved rapidly.

When the authors injected mouse fibroblasts lacking CAV1 and breast-cancer cells into mice, tumours were

## COMMUNITY CHOICE

The most viewed papers in science

## MATERIALS

## Graphene textiles for energy storage

**HIGHLY READ**  
on pubs.acs.org  
in June

Porous textiles coated with atom-thick sheets of carbon called graphene could underpin cheap and long-lasting energy-storage systems.

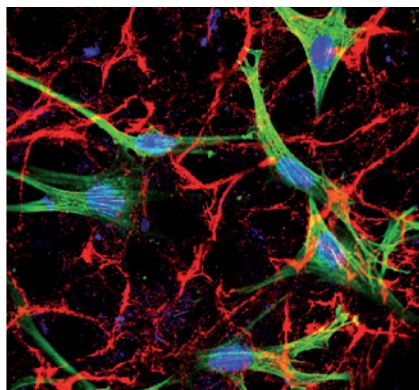
Zhenan Bao, Yi Cui and their colleagues at Stanford University in California dipped polyester fibres into a graphene solution and then deposited manganese dioxide onto the resulting structure. They used this as an electrode, combined with another made from carbon-nanotube-coated textiles, in a sodium sulphate solution. The resulting supercapacitor maintained a high level of energy storage and power delivery over 5,000 charge and discharge cycles, which is unusually long-lasting for manganese-dioxide-based electrodes. The fibres' three-dimensional porous structure has a larger surface area than conventional electrodes, enhancing performance.

Moreover, the system is made from abundant and environmentally friendly materials using a scalable process, the researchers say.

**Nano Lett.** doi:10.1021/nl2013828 (2011)

minimally invasive. But when they used CAV1-expressing fibroblasts, tumours grew more rapidly and invaded multiple organs.

**Cell** doi:10.1016/j.cell.2011.05.040 (2011)



## PALAEOLOGY

## Bridging the dino gap

Did an asteroid impact end the dinosaurs' reign? The controversy surrounding this question has been driven by a lack of dinosaur fossils dating to the period leading up to the impact in question.

Now researchers show that a fossilized dinosaur horn found in Montana is from the relevant period, suggesting that dinosaurs were not extinct before the impact.

Tyler Lyson at Yale University in New Haven, Connecticut, and his colleagues found a 45-centimetre-long ceratopsian brow horn 13 cm below the Cretaceous-Tertiary boundary, a geological feature thought to mark the time of the extraterrestrial impact in what is now Mexico around 65 million years ago. They identified the boundary through analysis of nearby rocks and Cretaceous fossils. The horn is the youngest non-avian dinosaur fossil yet discovered.

**Biol. Lett.** doi:10.1098/rsbl.2011.0470 (2011)

For a longer story on this research, see [go.nature.com/kx56ec](http://go.nature.com/kx56ec)

**NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# NEWS IN FOCUS

**PHILANTHROPY** Charities seek tangible returns from drug support **p.275**

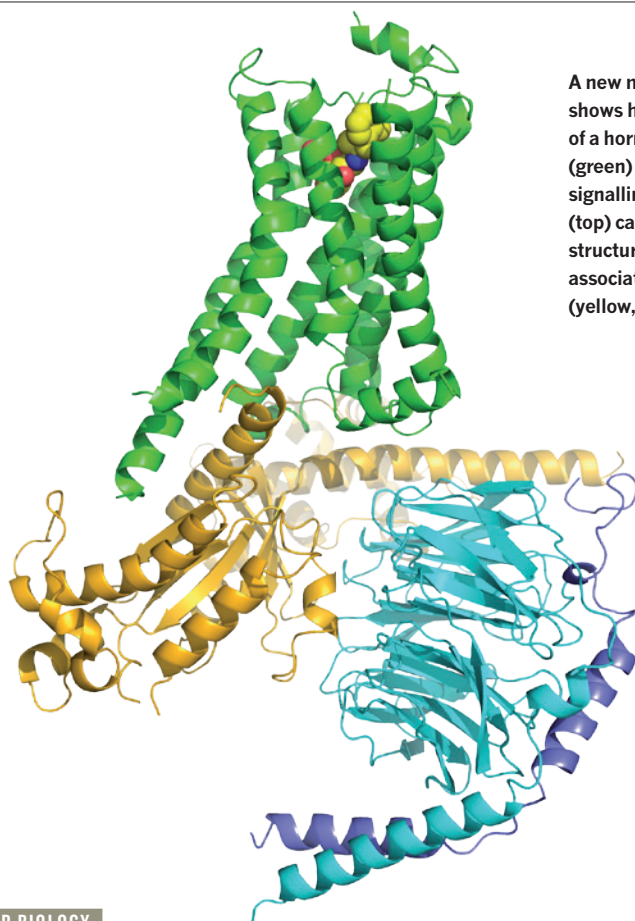
**NASA** Funding woes may ground flagship space telescopes **p.276**

**GENOMICS** Chips promise faster, cheaper DNA sequencing **p.278**



**SPORT** A new bid to make cycling drug-free **p.283**

REF. 1



A new molecular portrait shows how the activation of a hormone receptor (green) by a small signalling molecule (top) causes a dramatic structural shift in its associated G protein (yellow, blue and mauve).

## MOLECULAR BIOLOGY

# Cell signalling caught in the act

*Receptor imaged in embrace with its G protein.*

BY LIZZIE BUCHEN

Brian Kobilka knew that his postdocs didn't like him peeking at their experiments until they were finished. But he couldn't resist a quick look — after all, he and his entire field had been waiting for this result for more than 20 years.

As Kobilka peered through the microscope, the dream finally came into focus. Nestled in a drop of viscous liquid were tiny crystals, each trapping millions of copies of a fragile protein

complex. The structure of this complex could finally reveal how one of biology's most important signalling mechanisms, G-protein-coupled receptors (GPCRs), do their job. This structure, published online in *Nature*<sup>1</sup> by a team led by Kobilka at Stanford University in California and Roger Sunahara at the University of Michigan in Ann Arbor, now reveals the complete three-dimensional atomic structure of an activated GPCR — the  $\beta_2$  adrenergic receptor ( $\beta_2$ AR) — in a complex with its G protein.

GPCRs sit in the membranes of cells

throughout the body, where they detect signals from the outside world — such as light, odours and flavours — and signals from within the body, such as hormones and neurotransmitters. These signals are transmitted to the inside of the cell where they activate intracellular G proteins, which then trigger a variety of biochemical pathways.

The  $\beta_2$ AR is activated by the hormones adrenaline and noradrenaline, and kicks off the body's fight-or-flight response by speeding up the heart and opening airways. It is a key target for anti-asthma drugs. Kobilka's X-ray crystallographic snapshot of  $\beta_2$ AR associated with its G protein reveals some surprises, and could help in the design of more effective medicines — GPCRs are targeted by between one-third and one-half of all drugs on the market, including most of the best-sellers.

Before any protein can be imaged, it has to be crystallized. That is notoriously difficult for GPCRs, which need to be coaxed out of the cell membrane and kept stable in a fatty medium. The structure of the light-detecting GPCR rhodopsin was worked out in 2000 (ref. 2), but the GPCRs activated by hormones and neurotransmitters proved more intransigent. The first of these 'ligand-activated' GPCRs to yield to crystallization was  $\beta_2$ AR, which didn't give up its structural secrets until 2007, after decades of effort by Kobilka's group and others<sup>3–5</sup>. That opened the floodgates: the crystallographic structures of four other GPCRs have been solved in the past year<sup>6–9</sup>.

But understanding how GPCRs relay their signal meant crystallizing a complex of a receptor coupled to a G protein, an even harder task. The G protein, made up of three different subunits, is prone to detaching from the receptor and breaking apart, and the complex is about twice the size of  $\beta_2$ AR alone. Getting the structure of the  $\beta_2$ AR–G protein complex entailed developing new techniques to purify and stabilize it, including binding it to an antibody, and the testing of thousands of different crystallization conditions.

"This is a real breakthrough paper," says biochemist Stephen Sprang at the University of Montana in Missoula. "For a long time, many folks in the field have considered this the hoped-for structure that would ultimately provide a real understanding of how the receptors actually work."

Krzysztof Palczewski at Case Western Reserve University in Cleveland, Ohio, ►

► who was the first to crystallize rhodopsin<sup>2</sup>, agrees that the work is “a tremendous accomplishment”. But he is concerned that the engineered and antibody-stabilized proteins used in Kobilka’s study might not be a perfect match for the structure found in nature. Kobilka, however, says that his functional assays show that the engineered proteins behave like the natural proteins.

Researchers already knew that inactive G proteins are bound to a molecule of guanosine diphosphate (GDP) — a complex that Sunahara likens to a Pac-Man with something in its mouth. When a GPCR receives a signal, the receptor forces the G protein to spit out the GDP, allowing a molecule of guanosine triphosphate to swoop in and switch the G protein on.

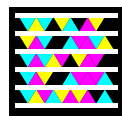
The structure now reveals how the activated receptor contorts to make this happen. Most surprisingly, it also shows that the G protein’s mouth plays wide open when the GDP departs. X-ray crystallography provides static images, so the exact sequence of events is unclear. “But now that we know it happens, it’s something we can study,” says Kobilka.

The discovery could provide unexpected clues to the molecular mechanism of the cholera toxin. The toxin forces G proteins to stay on all the time and continuously activate signalling pathways in intestinal cells. The affected cells release much of their water, leading to diarrhoea and vomiting. But the site that the toxin modifies is buried deep inside the G protein, which was “sort of puzzling”, says Sunahara. “How does it get to that buried site? Our structure showed us that the Pac-Man opens wide enough that it exposes the site.

And if that’s the way cholera works, it’s probably the way a lot of things interact with G proteins.”

“Brian’s struggled for this for such a long time,” says structural biologist Tracy Handel at the University of California in San Diego. “Thank God he got it, because, boy, he deserved it.” ■

➔ **NATURE.COM**  
Listen to more about  
this story on the  
Nature podcast  
[go.nature.com/nfqg2j](http://go.nature.com/nfqg2j)



Scan the tag above with the  
free app from [gettag.mobi](http://gettag.mobi)

1. Rasmussen, S. G. F. *et al. Nature* <http://dx.doi.org/10.1038/nature10361> (2011).
2. Palczewski, K. *et al. Science* **289**, 739–745 (2000).
3. Rasmussen, S. G. F. *et al. Nature* **450**, 383–387 (2007).
4. Rosenbaum, D. M. *et al. Science* **318**, 1266–1273 (2007).
5. Cherezov, V. *et al. Science* **318**, 1258–1265 (2007).
6. Shimamura, T. *et al. Nature* **475**, 65–70 (2011).
7. Xu, F. *et al. Science* **332**, 322–327 (2011).
8. Wu, B. *et al. Science* **330**, 1066–1071 (2010).
9. Chien, E. Y. *et al. Science* **330**, 1091–1095 (2010).



SCOTTS MIRACLE-GRO COMPANY

Glyphosate-resistant Kentucky bluegrass has outgrown US rules on genetically modified crops.

#### BIOTECHNOLOGY

## Transgenic grass skirts regulators

*Technological advances remove basis for government oversight of genetically modified crops.*

BY HEIDI LEDFORD

When the US Department of Agriculture (USDA) announced this month that it did not have the authority to oversee a new variety of genetically modified (GM) Kentucky bluegrass, it exposed a serious weakness in the regulations governing GM crops. These are based not on a plant’s GM nature but on the techniques used for its genetic modification. With changing technologies, the department says that it lacks the authority to regulate newly created transgenic crops.

The grass, a GM variety of *Poa pratensis*, is still in the early stages of development by Scotts Miracle-Gro, a lawn-care company based in Marysville, Ohio. The grass has been genetically altered to tolerate the herbicide glyphosate, which would make it easier to keep a lawn weed-free. On 1 July, secretary of agriculture Tom Vilsack wrote to the company to say that the variety “is not subject” to the same regulations that govern other GM crops. The decision allows Scotts to bypass the years of environmental testing and consultation

typically required by the regulators for GM plants, although the company says there are no plans to market this particular variety.

The grass can evade control because the regulations for GM plants derive from the Federal Plant Pest Act, a decades-old law intended to safeguard against plant pathogens from overseas. Previous types of GM plants are covered because they were made using plant pathogens. The bacterium *Agrobacterium tumefaciens* — which can cause tumours on plants — shuttled foreign genes into plant genomes. Developers then used genetic control elements derived from pathogenic plant viruses such as the cauliflower mosaic virus to switch on the genes.

By revealing similar elements in plants’ DNA, genome sequencing has liberated developers from having to borrow the viral sequences. And *Agrobacterium* is not essential either; foreign genes can be fired into plant cells on metal particles shot from a ‘gene gun’. Scotts took advantage of both techniques to construct the herbicide-resistant Kentucky bluegrass that put the USDA’s regulatory powers to the test.



"The Plant Pest Act was completely inappropriate for regulating biotech crops, but the USDA jury-rigged it," says Bill Freese, science-policy analyst at the Center for Food Safety in Washington DC. "Now we can foresee this loophole getting wider and wider as companies turn more to plants and away from bacteria and other plant-pest organisms." The USDA has not made public any plans to close the loophole and has also indicated that it will not broaden its definition of noxious weeds, a class of plants that falls under its regulatory purview, to facilitate the regulation of GM crops.

Nevertheless, *Agrobacterium* is still industry's tool of choice for shuttling in foreign genes, says Johan Botterman, head of product research at Bayer BioScience in Ghent, Belgium. The technique is well established for many crops, and particle bombardment is less predictable, often yielding multiple, fragmented insertions of the new gene.

But *Agrobacterium* isn't suitable for some new techniques. Many companies are developing 'mini-chromosomes' that can function in a plant cell without needing to be integrated into the plant's genome. Last summer, agribusiness giant Syngenta, based in Basel, Switzerland, conducted the first field trials of maize (corn) containing engineered mini-chromosomes, and showed that the mini-chromosomes, which carried multiple genes for insect and herbicide resistance, were stable in the field. "I would expect that by the end of the decade, this technology will be well used by many as a way to deliver large stacks of genes to plants," says Roger Kemble, head of technology scouting for Syngenta.

Other techniques under development insert foreign genes into designated sites in the genome, unlike the near-random scattering generated by *Agrobacterium*. In 2009, researchers at Dow AgroSciences in Indianapolis, Indiana, and Sangamo BioSciences in Richmond, California, announced that they had used enzymes called zinc-finger nucleases to insert a gene for herbicide resistance at a specific site in the maize genome (V. K. Shukla *et al.* *Nature* **459**, 437–441; 2009). Bayer is interested in harnessing other enzymes called 'meganucleases' to do the same type of targeted engineering, a strategy that Botterman says may make it possible to introduce multiple new traits into existing GM crops.

Regulators need to adapt to these new techniques, or run the risk of over- or under-regulating GM plants, says Roger Beachy, a plant biologist at Washington University in St Louis, Missouri, and former head of the USDA's National Institute for Food and Agriculture. The Kentucky bluegrass decision drives this point home, he says: "It really speaks to the importance of reviewing the regulatory process periodically to ensure that it is keeping up with the advances in technology." ■ **SEE EDITORIAL**

PAGE 265

## PHILANTHROPY

# Charities seek cut of drug royalties

*Non-profits that support medical research are angling for a share of the proceeds and intellectual-property rights.*

BY HEIDI LEDFORD

Early next year, a drug for cystic fibrosis is expected to come before the US Food and Drug Administration for approval. It is a moment that the Cystic Fibrosis Foundation (CFF) will have waited 12 years and invested US\$75 million to witness. Approval of the drug, VX-770 — developed by Vertex Pharmaceuticals of Cambridge, Massachusetts, with support from the foundation — would provide a new treatment for patients, and a revenue stream for the charity.

The CFF, based in Bethesda, Maryland, has a stake in the intellectual property underlying VX-770, and is entitled to royalties from sales of the drug. Such 'venture philanthropy' is increasing among charities. Like venture capitalists, non-profit groups are managing research projects, making funding dependent on the projects reaching predetermined milestones and potentially reaping a financial return. They are also keeping control over the fruits of their investment in case the journey from lab to treatment encounters obstacles.

"Philanthropies are looking to have more of a hand in managing intellectual property," says Timothy Coetzee, chief research officer of the National Multiple Sclerosis Society in New York, and former president of Fast Forward, the society's venture-philanthropy arm. Philanthropic donations for medical research are increasing (see 'Growing influence'), even as government granting agencies tighten their purse strings and venture capitalists cut back on biotechnology investments. As a result, non-profits have more bargaining power than ever before — especially for early-stage, high-risk projects that tend to be unattractive to private and federal investors.

"The charities are providing funds at the time when the risk is the very highest," says Ken Schaner, an attorney at Schaner & Lubitz — a law firm in Bethesda, Maryland, that specializes in working with non-profit

organizations. "But yes, they expect a return." The CFF is not alone: charities including the ALS Association in Washington DC, the Muscular Dystrophy Association in Tucson, Arizona, and the Wellcome Trust in London have also demanded royalties from some projects. Schaner says that the value of the return often depends on the size of the investment — for example, a foundation might be entitled to six times its input. In some cases, Schaner estimates that the payout could be as much as \$1 billion.

But organizations aren't interested only in generating revenue for their charitable work. Their involvement also helps to ensure that therapies reach the people who need them, in case anything happens to the drug companies with which they are collaborating.

In 2000, Schaner worked with the CFF to carve out a deal with Aurora Biosciences in San Diego, California — a pharmaceutical company that was later sold to Vertex — to develop the drug that was to become VX-770. The deal was one of the first examples of venture philanthropy.

But Schaner says that he couldn't sleep the night after the deal was signed. "I started thinking about what would happen if Aurora lost interest in the project. It could just sit there on the shelf untouched," he says. So he created an 'interruption licence' that is now used widely to give charities the intellectual-property rights behind a project if a company abandons it. ▶

**"The charities are providing funds at the time when the risk is the very highest. But yes, they expect a return."**

## GROWING INFLUENCE

Donations from charities to US biomedical research have tripled in the past decade.



SOURCE: FOUNDATION CENTER



► Those rights came in handy in another deal. The CFF had invested about \$25 million in a recombinant enzyme that could treat pancreatic deficiencies in people with cystic fibrosis. When the developer, Altus Pharmaceuticals in Waltham, Massachusetts, confessed that it could not afford a phase III clinical trial, the foundation snatched up the licence to the patent and shopped around for a new taker.

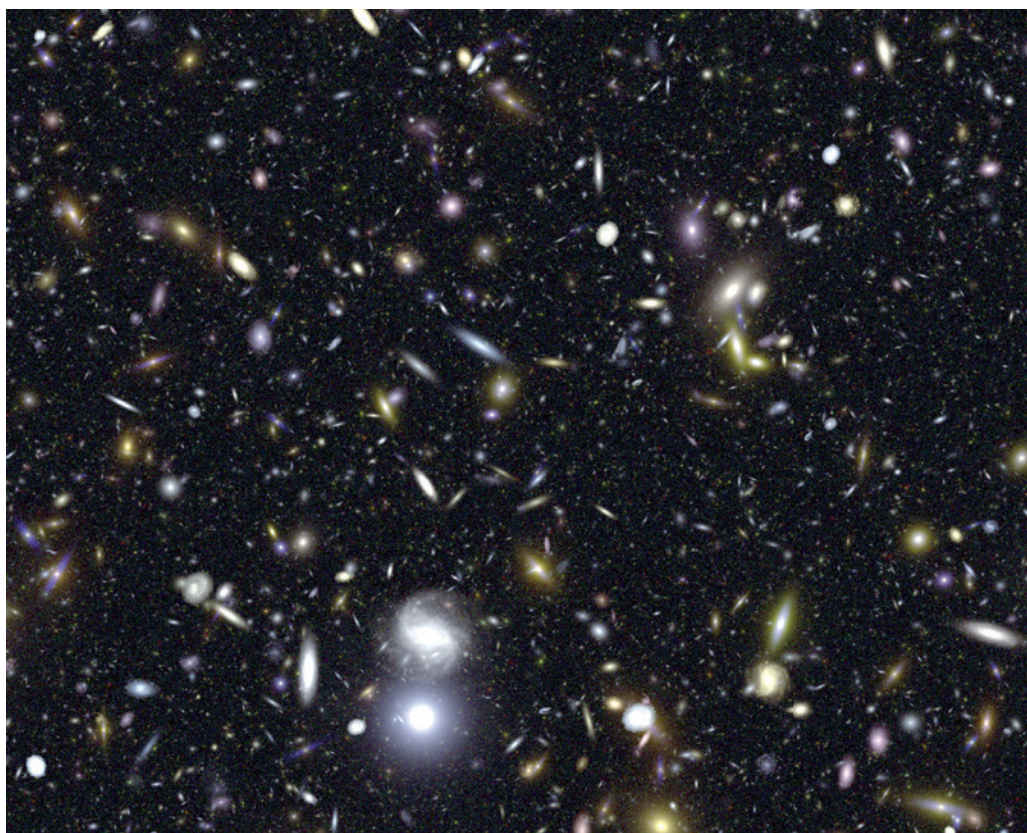
The technology ended up with Eli Lilly, a drug firm based in Indianapolis, Indiana. The foundation then sold off its royalty rights, funnelling the money into another programme. The recombinant enzyme came up for approval this year, but the Food and Drug Administration has requested further clinical trials.

Philanthropic organizations don't always go unchallenged: universities and companies can chafe at handing over intellectual-property rights. "Some philanthropies are getting more aggressive and greedy," says Jeffrey Quillen, a lawyer at the law firm Foley & Hoag in Boston, Massachusetts, who represents start-up companies and university spin-outs. "They see what big pharma gets from these deals and they decide they want stock or co-ownership of intellectual property, too." Some non-profits reduce their intellectual-property demands to ensure that the project doesn't stall because of disputes.

There is also strife when it comes to sharing royalties. "It's tough, but we'll do it sometimes," says Lita Nelsen, director of technology licensing at Massachusetts Institute of Technology in Cambridge. For example, the university might agree "if the foundation shares in the patent costs". Charities, for their part, tend to resist compensating universities for the 'indirect costs' that might result from a grant — which range from utilities to administrative support. That, notes Nelsen, adds to frustration in negotiations. "They think they're giving us money, but they're costing us," she says.

For the charities, royalties can help to fill the void left by the economic crisis. "Traditional fund-raising is still down for us," says Robert Beall, president of the CFF. "We took the Lilly royalties and put them right back into research — that's what we intend to do to make up for the deficit." The foundation reported more than \$53 million in royalty revenues last year.

But despite growing awareness of the importance of royalties and intellectual property, Schaner says that some non-profit organizations still give the issue short shrift. "Often, charities don't think past the first year or two when the grant is being made," he says. "They're so accustomed to clinical failures that they don't put enough emphasis on, 'We might have a success, and what happens then?'" ■ SEE EDITORIAL P.266



A simulated deep-field image of galaxies like those the James Webb Space Telescope might observe.

#### SPACE SCIENCE

# NASA telescopes face budget abyss

*Flagship missions at risk as astrophysics funding shrinks.*

BY ERIC HAND

As the space shuttle glides through its final week, another arm of the US space programme faces a bleak future. Astrophysics was once NASA's highest-funded science division and, with the Hubble Space Telescope, a long-time public-relations winner. But its two flagship telescope missions, ranked as the highest priorities for US astronomy, are now under threat as budget constraints start to bite.

Stung by spiralling costs and charges of mismanagement, the James Webb Space Telescope (JWST) — Hubble's long-awaited successor — is now seen by some critics as too expensive to fly. And the Wide-Field Infrared Survey Telescope (WFIRST), which would hunt for exoplanets and probe the poorly understood phenomenon known as dark energy, may take too long to develop to be

worthwhile. Added to that, the astrophysics division is facing a budget crunch while other science divisions within the agency weather the fiscal storm and even come out ahead.

"Clearly there's strong support for science," astrophysics director Jon Morse said at an advisory panel meeting on 13 July as he reviewed his division's place in the scientific pecking order at NASA. "The change here is about priorities."

With support from President Barack Obama, the agency's Earth science budget is at an all-time high. Over the next four months, the planetary science division is due to launch three major missions: to the Moon, to Mars and to Jupiter. And the heliophysics division plans to send a probe plunging into the blistering atmosphere of the Sun, closer than ever before. But because the overall NASA science budget is relatively flat, something had to give. Since 2008, astrophysics funding has plunged





NASA relative to other NASA science (see 'Falling fortunes') — and relative to physics and astronomy funding at other agencies.

"We're the orphan of the agency," says Alan Boss, chairman of the advisory panel and an astronomer at the Carnegie Institution for Science in Washington DC.

In Congress, the division faces outright hostility. While Boss lamented the budget trends at last week's meeting, a House appropriations committee was endorsing a bill that would cancel funding for the JWST. By all accounts, the 6.5-metre telescope will be at least as important to astronomy as Hubble has been. Designed to operate in the infrared, where the oldest celestial objects shine, the JWST could peer back to the Universe's first stars. It could also use the exquisite resolution offered by the vacuum of space to spot Earth-like planets. "Every once in a while, NASA does something that changes the game," says Michael Turner, director of the Kavli Institute for Cosmological Physics at the University of

Chicago, Illinois. "JWST is in that category."

But the cost of the JWST has also changed the game. In November, an independent panel called in by Congress blasted the project for mismanagement. It found that the telescope's price tag had ballooned to US\$6.5 billion and that its launch date would have to be delayed to 2015 (see *Nature* **468**, 353–354; 2010).

Even that projection seems to have been too optimistic. At the advisory panel meeting, Rick Howard, the recently installed JWST programme director, said that following his overhaul of the mission, the telescope will now launch in 2018 at the earliest.

"Of course people are disappointed," says project scientist John Mather of the Goddard Space Flight Center in Greenbelt, Maryland, which is managing the project. "I wanted it sooner, too." Although the extra time should be enough to resolve technical setbacks that have slowed the project, it also raises costs, as engineers and scientists must be employed for longer. Howard would not reveal exactly what the new price tag was, but he acknowledged that it would be more than \$6.5 billion. About \$3 billion has been spent already.

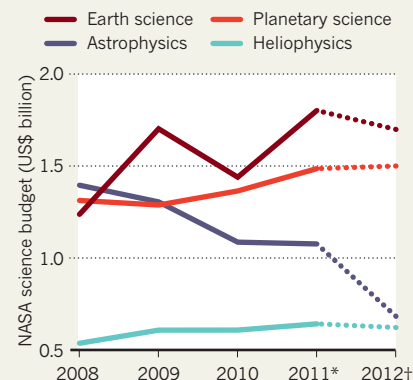
To get the mission back on track, NASA replaced all of the JWST's senior managers and put Howard at the helm. So far, he says, all of the telescope's 18 mirror segments have been polished and assembled. Engineers have also determined why some of the telescope's infrared detectors have begun to degrade. The bad detectors are being replaced at an additional cost of \$40 million to \$50 million, Howard says.

Many expect that the project's political defenders, such as Senator Barbara Mikulski (Democrat, Maryland), will be able to fend off the immediate threat in Congress. Garth Illingworth, an astronomer at the University of California, Santa Cruz, who was on the independent panel that reviewed the JWST, says that his real worry is whether the telescope will receive the massive infusion of funds it needs to get off the ground by 2018. The White House is currently requesting \$355 million for the project in 2012, and slightly more each year for the next four years. At those levels, Illingworth says, the telescope will never launch.

**"We're certainly more vulnerable than ever."**

## FALLING FORTUNES

Cuts to NASA's astrophysics division would see it fall from being the highest-funded science division to one of the least funded within the agency.



\*From the presidential budget request.

†From a US House of Representatives spending bill, which may change when reconciled with the Senate.

If the JWST is delayed yet again it will further imperil WFIRST, the next big astrophysics mission. The project, declared as the top priority in last year's decadal review of US astronomy by the National Academy of Sciences, is designed to measure the effects of dark energy — thought to be accelerating the expansion of the Universe — and to monitor distant stars for signs of exoplanets. But the rising cost of the JWST, together with NASA's declining astrophysics budget, could mean that the \$1.6-billion project might not get off the drawing board, let alone the launch pad.

The progress of a European Space Agency proposal to launch a similar telescope, called Euclid, by 2017 puts WFIRST in an even bigger bind. The proposal, submitted on 14 July, will face a final selection round in October. "If Euclid happens and is as good as they say, then I'm not sure [WFIRST] makes sense," says Turner.

Morse maintains that the division's budget woes will not affect other, smaller astrophysics missions scheduled for the next few years, but with two flagship missions in jeopardy that is small consolation to many astrophysicists.

"We're certainly more vulnerable than ever," says Boss, who worries that the seriousness of the situation may be lost on those outside the astrophysics community. "Maybe people are saying, 'They've got Hubble; that's all they need.'" ■



### Q & A

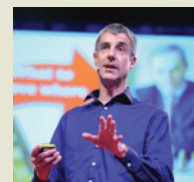


Michel Sidibé, executive director of UNAIDS, talks about China's HIV challenge. [go.nature.com/nph71d](http://go.nature.com/nph71d)

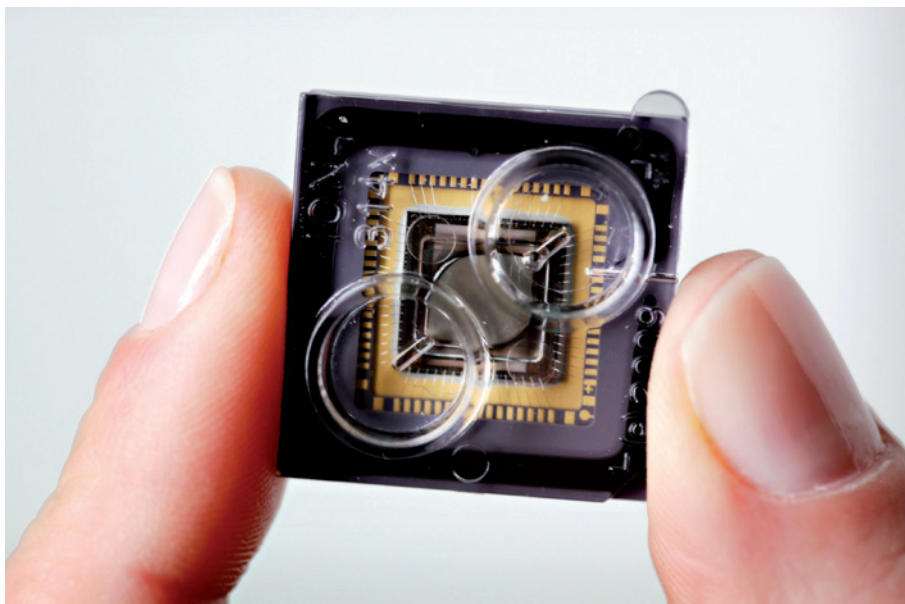
### MORE NEWS

- Pollutants' role in birth defects becomes clearer [go.nature.com/yldmjp](http://go.nature.com/yldmjp)
- Respiratory virus jumps from monkeys to humans [go.nature.com/7supdi](http://go.nature.com/7supdi)
- US deal will hurry hundreds of species onto endangered list [go.nature.com/4xbrcm](http://go.nature.com/4xbrcm)

### ON THE BLOG



TED conference: from the power of hugs to atheism 2.0 [go.nature.com/jt58q8](http://go.nature.com/jt58q8)



Each Ion Torrent chip sports 1.2 million DNA-testing wells.

# GENOMICS

# Chip chips away at the cost of a genome

*Ion-sensing method offers cheap sequencing in record time.*

BY GWYNETH DICKEY ZAKAIB

The latest contender in the race for the prized '\$1,000 genome' has proved its mettle in a singularly appropriate way: by sequencing the genome of computer pioneer Gordon Moore.

Like the computer chips made by Intel, the company that Moore co-founded, the Ion Personal Genome Machine (PGM) exploits semiconductor technology, with its ability to deliver ever-increasing speed and lower costs — a trend predicted by 'Moore's law' some 50 years ago. When Ion Torrent of Guilford, Connecticut, part of Life Technologies in Carlsbad, California, introduced the device late last year<sup>1</sup>, some scientists wondered whether it could live up to its promise to put a sequencer within the reach of any reasonably funded lab. Their doubts are likely to wane in the wake of the company's latest demonstration, published this week in *Nature* (see page 348).

In addition to producing a rough draft of Moore's genome, Ion Torrent has shown that its US\$49,500 device can read a bacterial genome in as little as two hours. "It's a quantum leap in terms of the time it takes to do an experiment," says Stephan Schuster, a molecular biologist at Pennsylvania State University in University Park, who has been testing the technology

for nearly a year. He has already published a paper that uses ion-sequencing data to investigate a cancer that is spreading rapidly among Tasmanian devils<sup>2</sup>. The PGM was also the first platform to unravel the genome of the *Escherichia coli* strain that began wreaking havoc in Germany in May<sup>3</sup>, delivering a sequence in just three days.

What makes the technology so quick and inexpensive is the novel way it detects the identity of the nucleotide bases in DNA. The Human Genome Project, which unveiled its landmark results a decade ago, relied on the laborious Sanger sequencing method. This involves building complementary DNA strands to match the original sample, until nucleotides labelled with a fluorescent dye are added to halt the process. The copied fragments are then sorted by size to determine the sequence of the original strand.

More recently, faster 'next generation' techniques were developed to read a DNA sequence by tracking the construction of a complementary strand as it actually happens. Most methods use fluorescent labelling to identify individual nucleotides as they are added. But these reagents are expensive — each sequencing run can

cost thousands of dollars, and may still take more than a week to complete.

Ion Torrent's device instead uses cheaper, natural nucleotides, and senses the hydrogen ions (protons) that are released as each nucleotide is incorporated onto the complementary DNA. "We made an array that literally sees chemistry," says molecular biologist Jonathan Rothberg, chief executive of Ion Torrent.

Microscopic beads carrying fragments of DNA are first loaded into 1.2 million 3.5-micrometre-wide wells covering a small chip that cost \$99. The chip is then flooded with washes of different nucleotides bearing the four bases that make up DNA, one after another. The wells are cleaned between each wash. If a nucleotide is complementary to the next unpaired base on the bead, it binds and gives off a hydrogen ion, changing the pH inside the well. This produces an electrical signal, indicating that the base in that particular wash is the next letter of the sequence. Each step takes less than five seconds, enabling a single chip to read about 25 million bases in a single two-hour run, and for just a few hundred dollars.

The technology's utility will ultimately depend not only on its cost, but also its accuracy, says Stephen Chanock, head of the Laboratory of Translational Genomics at the National Cancer Institute in Bethesda, Maryland. And, on that front, the PGM is still no match for the biggest, most expensive machines. Costing hundreds of thousands of dollars, they can read hundreds of billions of base pairs in a single run, and they are currently a more appropriate choice for tackling whole human genomes with high fidelity. Ion-chip sequencing is better suited to achieving fast results in smaller-scale projects, such as sequencing bacterial genomes or characterizing diseases by reading certain gene regions across many patients.

But Rothberg emphasizes that as transistors are packed more densely onto a single chip, the technology will become much more powerful.

**"It's a quantum leap in terms of the time it takes to do an experiment."**

Moore's genome required 1,000 ion chips — totalling 1 billion sensors — working in parallel. But the company is already testing an 11-million-well chip that could shrink that requirement tenfold and cut costs even further.

By switching to a manufacturing process able to create smaller features on a chip, Rothberg says, "we're very comfortable that we'll get way below a \$1,000 genome". ■

1. Katsnelson, A. *Nature* <http://dx.doi.org/10.1038/news.2010.674> (2010).
2. Miller, W. *et al. Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1102838108> (2011).
3. Turner, M. *Nature* <http://dx.doi.org/10.1038/news.2011.345> (2011).

ION TORRENT



## GENETICS

# How to build a better mouse

*The Collaborative Cross project will boost diversity and help the hunt for disease genes.*

BY EWEN CALLAWAY

It has taken nearly a century, but mouse geneticists are finally finishing the work started by Abbie Lathrop. The former schoolteacher from Massachusetts bred many of what became the first laboratory strains of mice in the early 1900s, yet her animals carried only a sliver of the genetic diversity found in wild mice. The hundreds of strains of laboratory mice used today still have a pretty narrow range of traits, which hampers the search for disease-causing genes.

Now, the Collaborative Cross, an ambitious project to create hundreds more mouse varieties representing a wider range of genetic diversity, is beginning to deliver its first animals. The new mouse strains have some very visible differences from one another — from variations in fur colour to tail length — and are already yielding clues to genes that help fend off fungal infection, which might not have been easily uncovered with standard lab strains<sup>1–3</sup>.

Many classic laboratory strains, such as C57BL/6 — the first mouse to have its genome sequenced — owe much of their genetic make-up to the same handful of ancestors.

These strains differ from each other in certain ways, such as the ability to battle infection, but not nearly as much as do wild mice. Huge chunks of the genomes of these strains are essentially identical, making it difficult and time-consuming to link particular traits to single genes within these genetic blind spots.

“Everyone realized there’s a truckload of variation that we aren’t seeing at all,” says Richard Mott, a statistical geneticist at the University of Oxford, UK, who is involved in the project.

Begun at the US Department of Energy’s Oak Ridge National Laboratory in 2005, the Collaborative Cross project selected five classic inbred strains, along with three more recently developed wild-derived strains, and began to breed them and their offspring together to reshuffle their genes.

To create genetically uniform inbred strains, brothers and sisters were mated for many generations. So far, the Collaborative Cross has established about 30 fully inbred mouse lines,

says Gary Churchill, a mouse geneticist at the Jackson Laboratory in Bar Harbor, Maine, one of the researchers who conceived the project.

The mice are already beginning to pay dividends. Fuad Iraqi, a geneticist participating in the Collaborative Cross at Tel Aviv University, Israel, tested 66 nearly inbred strains for their susceptibility to infection by

Rudolph Balling, director of the Luxembourg Centre for Systems Biomedicine, believes that the Collaborative Cross mice will become even more valuable when researchers start pooling their knowledge so that they can draw connections between seemingly distinct traits that have common genetic origins. “There has to be integrated database. That’s the key to the whole thing,” Balling says.

Steve Brown, director of the Mammalian Genetics Unit at MRC Harwell, UK, says the Collaborative Cross will mesh well with another project — the International Knockout Mouse Consortium — to create thousands of knockout strains collectively lacking nearly every mouse gene (see *Nature* **474**, 262–263; 2011). For instance, a gene knockout that affects a mouse’s sensitivity to diabetes could be linked to other traits of the syndrome, such as altered glucose metabolism, through the Collaborative Cross.

No database exists to help scientists forge such connections at present, and there is little capacity to shuttle hundreds of different mouse strains all over the world. So those involved with the Collaborative Cross are working with the University of North Carolina in Chapel Hill to

colonize the world’s labs with the new mice. They plan to send out breeding pairs of the first strains by the end of this year, with up to 100 strains available by 2012. “The idea is to make these available as broadly as possible,” Churchill says. ■



Mouse strains with greater genetic differences are ready to enter the lab.

*Aspergillus fumigatus*, a soil fungus that causes a respiratory disease in humans.

Depending on the strain, the mice survived between 4 and 28 days after infection. On the basis of genotype information for the new strains and the genome sequences of the eight founder strains, Iraqi’s team mapped these differences in survival time to just a handful of genomic regions, containing a small number of genes<sup>3</sup>. Future studies in ‘knockout’ mice lacking these genes should pin down exactly which ones are responsible for fungal resistance, Iraqi says.

Getting to this point with the Collaborative Cross mice took only a year, compared with the decade and a half Iraqi estimates it would have taken with the classic strains. “It is amazing,” he says. His team is taking the same approach to map genes involved in defence against the bacterium *Klebsiella pneumoniae* and other traits.

“I don’t think results are going to trickle out, they’re going to be bursting,” Churchill says.

1. Aylor, I. et al. *Genome Res.* <http://dx.doi.org/10.1101/gr.111310.110> (2011).
2. Philip, V. et al. *Genome Res.* <http://dx.doi.org/10.1101/gr.113886.110> (2011).
3. Durrant, C. et al. *Genome Res.* <http://dx.doi.org/10.1101/gr.118786.110> (2011).

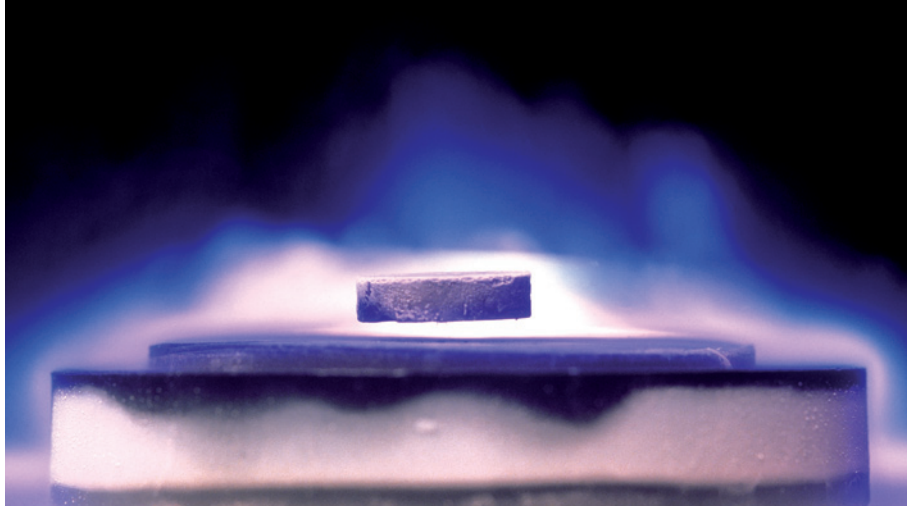
## CORRECTION

The News story ‘Paxil study under fire’ (*Nature* **475**, 153; 2011) gave the wrong affiliation for Charles Bowden. He is a clinical professor of psychiatry and pharmacology at the University of Texas Health Science Center, San Antonio.

# STILL IN SUSPENSE

A quarter of a century after the discovery of high-temperature superconductivity, there is still heated debate about how it works.

BY ADAM MANN



“Even bouncers in New York City night-clubs were aware of our notoriety,” says Paul Grant, thinking back to the 1987 March meeting of the American Physical Society (APS).

The hype had been building for months, as newspapers, magazines and morning television talk shows heralded jaw-dropping announcements from physics labs. A technological revolution seemed at hand, promising an era of levitating trains, coin-sized computers and power lines that could span continents without losing energy. When the meeting finally convened, says Grant, a physicist at the energy consulting firm W2AGZ Technologies in San Jose, California, anyone with an APS badge who arrived at a trendy club aptly named ‘The Limelight’ was ushered straight to the front of the queue.

Yet the public’s excitement was nothing compared with the eager frenzy of the physicists. On the evening of Wednesday 18 March, more than 1,800 APS attendees squeezed into a ballroom at the New York City Hilton (while another 2,000 milled outside) to watch a marathon set of presentations that lasted more than 7 hours. At the sometimes-raucous symposium — dubbed the ‘Woodstock of physics’ — researchers devoured the latest findings on what was easily the most astonishing discovery their field had seen in a generation: materials that became superconductors at high temperatures.

‘High-temperature’ was a relative term: even the best of the materials would not transition to become superconducting — having no resistance to an electric current — until it was chilled below 93 K (roughly 200 °C below room temperature). But that was nearly four times higher than the transition temperature of any previously observed superconducting material, and shattered what had once seemed to be a solid theoretical upper limit of 30 K. Everyone in the

A sample of a high-temperature superconductor hovers in a magnetic field.

## 1911 ▶

### A century of superconductivity

Heike Kamerlingh Onnes (seated centre front) and his colleagues discover superconductivity. He receives the Nobel prize in 1913.



## 1957



John Bardeen, Leon Cooper and Robert Schrieffer (left to right) publish a theory of superconductivity that predicts a maximum transition temperature of 30 K. They are awarded the Nobel prize in 1972.

## 1986

Georg Bednorz (left) and Alex Müller find a copper oxide material that becomes superconducting at 35 K.





ballroom knew that, whatever was going on, it was something profoundly new.

Better still, they knew that 93 K could be achieved easily with cheap, plentiful liquid nitrogen as a coolant, instead of the expensive, tricky-to-handle liquid helium required by the earlier superconductors. Suddenly, applications of superconductivity such as lossless power lines seemed economically feasible. And the room was alive with an even more electrifying idea: could there be materials that superconduct without any refrigeration at all?

But 25 years after the publication of the first paper on high-temperature superconductivity<sup>1</sup>, such materials remain a dream. So do most of the miraculous-sounding applications. And so does a deep understanding of what is going on. Despite increasingly refined experimental techniques and nearly 200,000 published papers, physicists still do not have a complete theoretical explanation for high-temperature superconductivity. “It’s not that there’s no theory; there are lots of theories — just none that most people agree on,” says John Tranquada, a physicist at the Brookhaven National Laboratory in Upton, New York.

## SLOW PROGRESS

Still, history offers some reassurance. Physicists took 50 years to understand conventional superconductivity — which was discovered 100 years ago in the laboratory of Heike Kamerlingh Onnes, at Leiden University in the Netherlands (see ‘A century of superconductivity’). On 8 April 1911, after testing for electrical resistance in a sample of mercury at 3 K, Onnes wrote down “*Kwik nagenoeg nul* (Mercury practically zero)”, marking the first observation of a superconductor.

A step towards an explanation of superconductivity came in the 1920s with the development of quantum mechanics, which provided an underlying model for the structure of ordinary metals. Metal atoms form a regular crystalline lattice and hang on to a tightly

bound inner core of electrons. But their loosely attached outer electrons become unbound, collecting into a mobile ‘electron sea’. Under the influence of an electric field, this ocean of free electrons will drift throughout the lattice, forming the basis of conductivity.

In a normal metal, this motion isn’t always predictable: no matter how cold it gets, random thermal fluctuations scatter the electrons, interrupting their forward motion and dissipating energy — thereby producing electrical resistance. But as some metals are cooled to temperatures close to absolute zero, the electrons

**“It’s not that there’s no theory; there are lots of theories — just none that most people agree on.”**

suddenly shift into a highly ordered state and travel collectively without deviating from their path. Below a critical temperature that is unique to each of these metals, the electrical resistance falls to zero and any current flows practically forever. They become superconductors.

But why does this ordered state form? In February 1957, three physicists — John Bardeen, Leon Cooper and Robert Schrieffer, all then at the University of Illinois in Urbana-Champaign — published the first complete answer<sup>2</sup>.

According to their proposal, now known as BCS theory, an electron moving through a positively charged lattice of atomic nuclei leaves behind a small wake, like the deformation caused by a bowling ball rolling across a mattress. The distortion pulls in another electron, and the two become what is known as a Cooper pair. If many such pairs form, as

happens at very low temperatures, their quantum-mechanical wavefunctions align, drawing the pairs into a collective state known as a condensate. Once there, they keep one another in check because breaking up one pair would raise the energies of all the others. The net result is that they all flow together without interruption, creating superconductivity.

The theory was very successful, making many predictions that were quickly confirmed by experiment. But it also implied that the forces binding the Cooper pairs were very feeble, so they would be ripped apart by thermal vibrations at anything other than extremely low temperatures. “Armies of researchers in the 1950s and ’60s worked on improving the temperature range,” says Jan Zaanen, a theoretical physicist at Leiden University. “But they soon realized that they could not give rise to superconductivity above 25 K or 30 K” — temperatures that generally require elaborate cooling systems for liquid helium, which boils at 4.2 K.

This did not stop the use of superconducting wires and films in certain high-value applications such as medical magnetic resonance imaging (MRI) machines and particle colliders. But the expense seemed to rule out any wider application.

Then, in June 1986, physicists Georg Bednorz and Alex Müller at the IBM Laboratory in Zurich, Switzerland, reported<sup>1</sup> that they had created a material that became superconducting at 35 K. The finding was dramatically confirmed in January 1987, when physicists in the United States found a material in the same class that became superconducting at 93 K (ref. 3). The Woodstock of physics followed barely two months later.

One of the many astonishing aspects of Bednorz and Müller’s work was that they were looking not at metals, but at insulating materials called copper oxides, which physicists would soon dub cuprates. In particular, they were investigating what happens when a cuprate is ‘doped’, or has foreign elements

# 1987

**January:** High-temperature superconductivity is confirmed in cuprates, this time at a temperature of 93 K.

**March:** The American Physical Society hosts the ‘Woodstock of physics’ (pictured). And Phillip Anderson posits the resonating-valence-bond theory as the mechanism for high-temperature superconductivity.



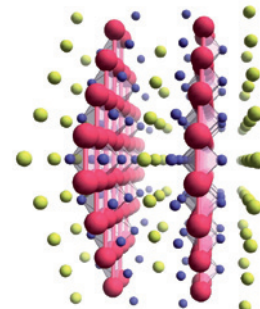
# 1991

Philippe Monthoux, Alexander Balatsky and David Pines publish the spin fluctuation theory of high-temperature superconductivity.

# 1993

Researchers discover a material that becomes superconducting at 135 K, setting a world record for the highest transition temperature.

# 2008



Hideo Hosono and his co-workers discover a new class of superconductors, iron pnictides (pictured).



such as lanthanum or barium introduced into the parallel planes of copper and oxygen that comprise its structure. What they found was that the foreign atoms freed up the outer electron of some of the copper atoms, which then flowed through the lattice. If the cuprate was then cooled enough — to a temperature that depended on how it was doped — the electrons would flow freely, and the material would become superconducting.

This strange state of affairs — superconductivity in an insulator — quickly led physicists to re-examine their basic ideas about condensed matter. But because some of the experiments were unknowingly done on impure samples, people were having trouble reproducing the results. “The first years of the field were very confusing,” says Patrick Lee, a physicist at the Massachusetts Institute of Technology in Cambridge. Hypotheses invoking bizarre and exotic physics cropped up, often without much evidence to back them up.

The field soon broke up into competing camps, each advocating a different theory. Researchers would often ignore data that did not jibe with their pet theory, clinging almost religiously to their ideas, and attacking those who believed otherwise.

Kathryn Moler, a physicist at Stanford University in California, recalls a colloquium in which a scientist in the audience stood up, pointed a finger at the speaker and shouted, “Liar! Liar! Ladies and gentlemen, that man is a liar — don’t listen to a word he’s saying!” Igor Mazin, a physicist at the Naval Research Laboratory in Washington DC, remembers a conference in 1989 when physicists promoting the different theories stood on stage “yelling like schoolchildren”.

Eventually, the cacophony sorted itself into the two theories with which most physicists now work. The first, resonating-valence-bond theory<sup>4</sup>, is largely the creation of Philip Anderson, a condensed-matter physicist at Princeton University in New Jersey. The theory states that the electron-pairing mechanism is imprinted in the cuprates’ structure. Neighbouring copper atoms can become linked through chemical valence bonds, in which they share electrons with opposite spins. Typically, the bonding locks these spin pairs in place, preventing any current from being carried. But when the material is doped, the pairs become mobile and the valence bonds become Cooper pairs that condense into a superconductor.

The second theory, called spin fluctuation<sup>5</sup>, has the most support in the community. Devised by Philippe Monthoux of the University of Edinburgh, UK, Alexander Balatsky from Los Alamos National Laboratory in New Mexico and David Pines from the University of Illinois–Urbana Champaign, it posits that without doping, cuprates are locked into an ordered state called an antiferromagnet. That means that the outer electron on each copper atom lines up such that its spin is opposite to

that of its neighbour: one electron will have its spin up, the next down, the next up, and so on. The magnetic fields produced by the spins lock the electrons in place. But in doped cuprates, the foreign atoms break up this rigid checkerboard pattern, giving the spins room to wobble. A passing electron can then set up a pulsating pattern of spins analogous to the lattice distortions of conventional superconductivity. This disturbance then draws moving electrons together, allowing them to associate into Cooper pairs and achieve a superconducting state.

In the early days, says Tranquada, advocates of these two mechanisms were at loggerheads as much as anyone else in this field. But after a while, he says, “it becomes easier to relax a little bit and try to start discussing where the points of agreement are and where the points of disagreement are. We can get beyond opinions and try to make some progress by agree-

## “Ladies and gentlemen, that man is a liar — don’t listen to a word he’s saying!”

ing on some experiments or calculations that may help.” Most researchers now broadly agree on many aspects, such as the importance of magnetic interaction.

Things have also calmed down a bit in the laboratory, as improved techniques have helped researchers to weed out the more exotic theories and refine those that remain. A good example is angle-resolved photoemission spectroscopy (ARPES), a method that uses high-energy photons to probe what electrons are doing. “In 1993, the best we could do was four spectra in 12 hours,” says Zhi-Xun Shen, a physicist at Stanford University who works with ARPES. “One of vastly superior quality now takes 3 seconds.”

And in 2008, Hideo Hosono and his colleagues at the Tokyo Institute of Technology in Japan discovered a second class of high-temperature superconducting material — this time based on iron and arsenic — called pnictides<sup>6</sup>. These materials superconduct at lower temperatures than most cuprates — often only below 40 K — but they have given theorists a new arena for testing their ideas.

“It’s almost like a do-over,” says Thomas Maier, a physicist at Oak Ridge National Laboratory in Tennessee. Pnictides have a more complex structure than cuprates, but they might help to uncover which phenomena are central to high-temperature superconductivity, and which are simply due to the copper oxide structure.

Moreover, finding the pnictides has reassured researchers that they might be able to find other high-temperature superconductors, providing more information or perhaps even a path to the elusive room-temperature superconductor. “Once there’s two, there’s a high probability of there being more,” says Andrew Millis, a physicist at Columbia University in New York.

Researchers have made progress in practical applications. In the past five years, for example, they have managed to string cuprate materials into superconducting tape that can be used in power transmission cables or MRI machines cooled with liquid nitrogen.

### THE ROOT OF THE MATTER

No one is predicting a full understanding of high-temperature superconductivity any time soon — not least because such an account would have to make sense of the huge number of papers. “A rich enough theory should explain everything and not just cherry pick,” says David Pines, a physicist from the University of Illinois at Urbana-Champaign.

But it’s not always clear exactly what needs to be explained. Roughly 15 years ago, for example, researchers discovered that some high-temperature superconductors allow electron pairs to form above the transition temperature. In this ‘pseudogap’ regime, the material spontaneously organizes itself into stripes: linear regions that act like rivers and carry electron pairs through the insulating landscape where electrons remain stuck in place. “It’s a precursor state to the superconducting state and is therefore fundamental to understanding this problem,” says Ali Yazdani, a physicist at Princeton University. Not so, says Pines, who thinks the pseudogap state “interferes with superconductivity but is not responsible for it”.

Much as physicists had to wait for highly developed quantum-mechanical tools to unlock the secret behind traditional superconductivity, researchers today may require future ideas to complete their task.

If nothing else, the field’s early quarrels have ensured that only the most determined researchers have stayed. Those remaining are perhaps humbled by their experiences. “I think our biggest problem has been human fallibility,” says Anderson. And perhaps these initial difficulties have helped to forge theories that can stand the test of time. “In the end, it’s your competitor that makes you strong,” says Shen. ■

**Adam Mann** is a freelance writer based in Oakland, California.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **64**, 189–193 (1986).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **106**, 162–164 (1957).
3. Wu, M. K. *et al.* *Phys. Rev. Lett.* **58**, 908–910 (1987).
4. Anderson, P. W. *Science* **235**, 1196–1198 (1987).
5. Monthoux, P., Balatsky, A. V. & Pines, D. *Phys. Rev. Lett.* **67**, 3448–3451 (1991).
6. Takahashi, H. *et al.* *Nature* **453**, 376–378 (2008).



# RACING JUST TO KEEP UP

*Anti-doping researchers are looking for new ways to catch cheaters. Can a biological passport help to save the sport?*

BY EWEN CALLAWAY

Cyclist Borut Božič drew his hands to his chest with a look of joy, disbelief and exhaustion after defeating some of the world's best sprinters in the Swiss village of Tobel. His stage victory at the week-long Tour de Suisse last month netted the 30-year-old Slovenian a €4,000 (US\$5,600) bonus and probably helped to secure his spot in this month's Tour de France, cycling's most prestigious race.

His stage win also automatically earned Božič a trip to a cramped medical trailer. Inside, he and three other riders each filled two small jars with urine. The containers were sealed, anonymized and sent to the Swiss Laboratory for Doping Analyses in Lausanne, where technicians would test them for traces of steroids, stimulants and a potent blood-boosting drug

called erythropoietin (EPO).

Such tests have become as much a part of professional cycle racing as carbon-fibre bicycles, but decades of doping scandals show that they are no guarantee of a drug-free race. It is tough to name a Tour de France win in recent years that has gone unmarred by doping accusations. Last year's winner, Alberto Contador, tested positive for the banned drug clenbuterol. He has successfully argued that it came from contaminated meat, but an arbitration hearing could still erase his victory. And last year, it was revealed that seven-time Tour de France winner Lance Armstrong has been the focus of a US Justice Department investigation into doping — although he has never been disciplined and maintains that he never doped. Confronted with increasingly sophisticated dopers, anti-doping scientists face a daunting game

of catch-up. "This is an endless whirl," says Martial Saugy, the director of the Lausanne laboratory.

In hopes of slowing the whirl, Saugy's team has pioneered a new kind of anti-doping test: the biological passport. Instead of scouring an athlete's urine for traces of drugs or their breakdown products — as the Lausanne lab would do for Božič's sample — the passport builds up a profile of an individual over time and tries to detect biochemical changes that might indicate doping.

Since 2008, Saugy's laboratory and the International Cycling Union (UCI), cycling's international federation based in Aigle, Switzerland, have created biological passports for hundreds of professional cyclists, some containing data from dozens of blood draws. Other sports are looking to follow suit. Some researchers say that the passport offers the best line of defence against EPO



use, which has bedevilled inspectors for the past two decades; and biological passports to detect steroid and growth-factor doping are in the works. The technology may see its Olympic debut at the games in London next year. Still, critics — and some athletes — say that it is no match for determined dopers.

"The biological passport is a joke," said Floyd Landis, a former US pro cyclist, to sport-news website ESPN.com in May 2010. After losing a costly four-year battle to overturn his conviction for using steroids during the 2006 Tour de France, Landis admitted to doping for much of his career and said that pro cyclists knew how to defeat the biological passport before it was introduced. But the passport has already led to convictions, and — perhaps briefly — shifted the advantage back to the testers. "I think we are forcing people to decrease their doping habits," says Saugy.

### AN ENDLESS CYCLE

Anti-doping efforts started in earnest after the 1960 Olympic Games in Rome. During a team time trial, 23-year-old Danish cyclist Knud Enemark Jensen collapsed, fractured his skull and died. An autopsy reportedly found traces of amphetamine and a blood-vessel dilator in his system. Although the drugs might not have caused his death, the episode forced cycling officials to take a closer look at doping. The UCI banned some performance enhancers, and in 1967 the International Olympic Committee established a commission to ferret out doping in sport.

The task is thankless: anti-doping agencies thwart one cheating strategy, only for another to emerge. The 1972 Olympic Games in Munich, Germany, ushered in testing for stimulants, but athletes had started to take anabolic steroids. A test for steroids arrived at

the next summer Olympics, in Montreal, Canada. But four years later, at the Moscow Olympiad, athletes had moved on to undetectable, naturally occurring hormones, such as testosterone. Anti-doping authorities now measure the ratio of testosterone in the blood to a related molecule called epitestosterone. In response, some athletes have reportedly found ways of regulating epitestosterone to keep the ratio in check.

For cycling and other endurance sports, human recombinant EPO fuelled a doping revolution. EPO is a natural hormone that promotes production of oxygen-carrying red blood cells. The first synthetic, or recombinant, version was developed by the biotechnology company Amgen in Thousand Oaks, California, and in 1989 it was approved by the US Food and Drug Administration to treat anaemia. It also offered cyclists an easy endurance boost that helped them to excel in gruelling stage races. The drug is nearly identical to the hormone naturally churned out by the kidneys, so was impossible to detect. It is also easier to administer than blood transfusions, which had been used to the same effect.

"In the 1990s and 2000s, it was quite easy for the cheaters to use huge amounts of EPO," says Saugy. Don Catlin, a pharmacologist who used to run an anti-doping laboratory at the University of California, Los Angeles, has a grimmer view. "Everyone in cycling was doping," he says.

Without a test for EPO, cycling regulators turned to an indirect measurement called the haematocrit — the percentage of blood volume made up of red blood cells. Typically, red blood cells account for 40–45% of the blood, but in the heyday of EPO doping, some riders were showing up at starting lines with haematocrits of more

than 60%. Their blood was so viscous that they would collapse before races, says Neil Robinson, who led the development of the passport at the Lausanne laboratory. The UCI instituted a 'no-start' rule, disqualifying riders if their haematocrits on the morning of a race were above 50% for men and 47% for women. So cyclists began diluting their EPO-boostered blood with saline solution to keep their haematocrits below the threshold, says Robinson.

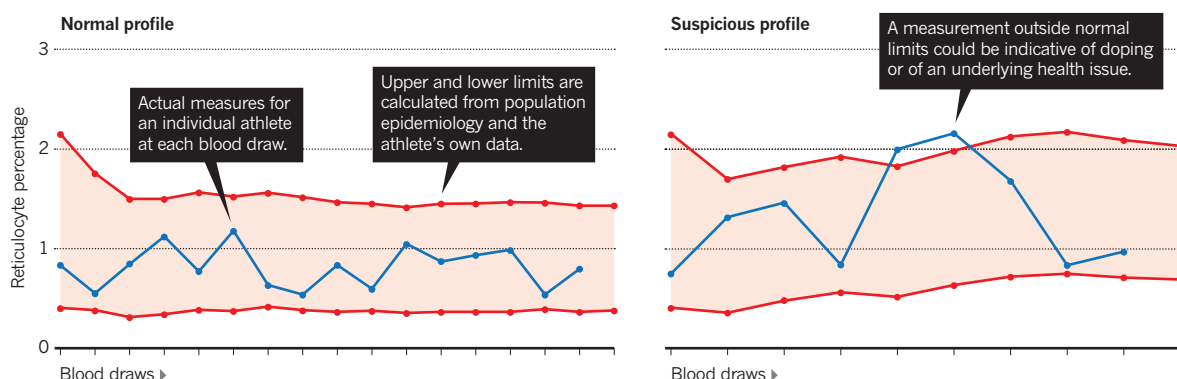
The drug companies that produce EPO have helped anti-doping laboratories to develop direct tests based on subtle biochemical differences between the recombinant molecules and the natural form. The first of these was approved for use in 2000. But athletes increasingly obtain knock-off forms produced in China and India, and researchers have struggled to keep up, says Robinson: "The solution is the passport."

The passport started taking shape in 1999, when Robinson and Saugy began clinical studies of EPO doping in volunteers. "We immediately realized that there were major differences between subjects," says Robinson. For example, in one volunteer, levels of immature red blood cells called reticulocytes might rocket up in response to the hormone, whereas in another, they might barely rise. The researchers realized that instead of comparing such blood metrics against a wide range of values based on the general population, it made more sense to use an athlete as his or her own control and look for unusual fluctuations.

Today, the passport is an electronic record of several different characteristics of red blood cells — haematocrit, the concentration of the blood protein haemoglobin, the percentage of reticulocytes and other metrics — collected

## GOOD BLOOD, BAD BLOOD

The biological passport tracks nine blood characteristics for an athlete over time. Below are normal-looking (left) and suspicious-looking (right) measurements for one of these: the percentage of reticulocytes, or immature red blood cells, in the blood. Although an abnormal result for one characteristic doesn't necessarily raise suspicion, abnormal readings for more than one could indicate that the athlete is doping.



periodically in and out of competition for an individual athlete (see 'Good blood, bad blood'). A statistical model that accounts for factors such as an athlete's sex or the altitude at which a sample was collected (thinner air boosts red-blood-cell production) estimates the probability that a rider's profile is anomalous. "The model will not tell you whether they've doped or not — it tells you the degree of abnormality," says Robinson.

### MODEL OF HONESTY

A panel of anti-doping experts reviews profiles identified as suspicious and determines which cases merit a full-blown investigation. Although it is generally used to target riders for direct testing, cyclists have been successfully prosecuted on the basis of their biological passports alone. And in March, the Court of Arbitration for Sport, an international supreme court of sport based in Lausanne, upheld two of these prosecutions, further legitimizing the approach. More cases may be on the way. A report leaked to the French sports newspaper *L'Équipe* revealed a list compiled by the UCI, rating last year's Tour de France riders on a scale of 0 to 10 on the basis of their biological passports. From a total of 198 riders, 42 were rated at 6 or higher, meaning that they showed "overwhelming" evidence of doping, according to *L'Équipe*. Although it isn't proof of doping, the list may be used in deciding which riders to scrutinize in the future.

"There are still chinks in the armour," says Catlin. A team led by Michael Ashenden, an anti-doping researcher who heads the Science and Industry Against Blood Doping consortium in Gold Coast, Australia, simulated EPO 'microdosing' in ten volunteers<sup>1</sup>. They received small intravenous injections twice weekly for 12 weeks. The treatment boosted the subjects' haemoglobin mass by 10%, equal to two bags of transfused blood, but the biological passport didn't flag a single profile as suspect.

In another study<sup>2</sup>, Carsten Lundby, a cardiac physiologist at the University of Zurich in Switzerland, and his team subjected three groups of volunteers to different EPO regimens for ten weeks. A testing approach similar to the biological passport caught only 58% of the doped volunteers. "I'm happy I'm not working in anti-doping, because it must be frustrating," says Lundby.

Some researchers say that the statistical model underpinning the passport might produce an unacceptably high

number of false positives — clean riders who look dirty on a test. Clifford Spiegelman, a statistician at Texas A&M University in College Station, complains that the model wrongly assumes that biological variations follow what statisticians call a normal distribution. Normal distributions resemble bell-shaped curves, with few outliers. The problem, says Spiegelman, is that biological measurements are chock

full of outliers — far more than would be predicted by a normal distribution. Proponents of the passports are "presenting themselves as more accurate than they really are", he says, and he estimates that the false-positive rate of the passport could be off by a factor of 10 or even 100.

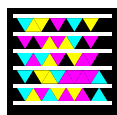
Pierre-Edouard Sottas, a Lausanne-based scientist with the World Anti-Doping Agency who developed the statistical model underpinning the passports, says that tests on thousands of clean athletes show that the blood characteristics used do follow a normal distribution. Moreover, he notes that a panel of experts, not his statistical model, makes the final decisions about an abnormal profile.

### NO SIGN OF THE FINISH LINE

Robinson acknowledges that the passport cannot catch everyone, but it could deter dopers. The UCI points to a study<sup>3</sup> from its scientists indicating that the incidence of blood metrics that suggest doping has declined since the introduction of the passport.

Anti-doping scientists think that they can improve the tests using tactics such as monitoring sudden spikes in performance, which could indicate something other than intensive training. Robinson and his team want to incorporate information garnered through police investigations of telephone and customs records into the biological passport's predictive model,

**ON NATURE.COM**  
To listen to a podcast about doping in sport, visit:  
[go.nature.com/nfq2j](http://go.nature.com/nfq2j)



Scan the tag above with the free app from **gettag.mobi**

so that suspicious behaviour and blood chemistry could both be used to flag a rider for closer follow-up. "We have to use the same approach as a crime scene," says Robinson.

His team is

also developing versions of the passport to detect steroid and growth-hormone abuse by charting changes in the urine or blood levels of compounds such as testosterone and insulin-like growth factor-1. These researchers and others are also looking to improve the biological passport by searching for new molecular indicators of blood doping. For example, according to unpublished research by the Lausanne

**"WE HAVE TO USE THE SAME APPROACH AS A CRIME SCENE."**

lab, circulating levels of a microRNA called miR-144, which is involved in regulating red-blood-cell production, spike after volunteers take EPO. Yorck Olaf Schumacher, an anti-doping scientist at the University of Freiburg in Germany, says that his lab has found changes in gene expression in response to transfusions of a patient's own blood, which can't be detected using conventional markers. Robinson says that it will be several years before these new markers make their way into the biological passports. "We need to validate all these approaches, and that gets tricky."

As the three-week, 3,400-kilometre trek of the Tour de France nears its finish on 24 July on the Champs-Élysées in Paris, Božič has yet to duplicate his Tour de Suisse stage victory. But his biological passport has gained another data point. Before setting off on this year's Tour de France, Božič and the other 197 riders gave blood samples for their passports, says Robinson, whose team plans to use these data anonymously to estimate the prevalence of blood doping in this year's race.

The team hopes that the passport will keep more riders honest. But after running an anti-doping laboratory for a quarter of a century, Catlin is convinced that tests, no matter how sophisticated, will never keep up with the most determined dopers. "For every move to the right, the other guys are moving to the left and it balances out again." ■

**Ewen Callaway** writes for Nature from London.

1. Ashenden, M., Gough, C. E., Garnham, A., Gore, J. C. & Sharpe, K. *Eur. J. Appl. Physiol.* <http://dx.doi.org/10.1007/s00421-011-1867-6> (2011).
2. Bornø, A., Aachmann-Andersen, N. J., Munch-Andersen, T., Hulston, C. J. & Lundby, C. *Eur. J. Appl. Physiol.* **109**, 537–543 (2010).
3. Zorzoli, M. & Rossi, F. *Drug Test. Anal.* **2**, 542–547 (2010).



# COMMENT

**CONSERVATION** Scientists must think bigger and faster to tackle today's problems **p.290**

**EPIDEMIOLOGY** History of disease maps shows some things never change **p.292**

**HISTORY** Spacesuits, where fashion met technology **p.294**

**FILM** In conversation with Richard Berendzen, on NASA and sci-fi **p.295**



S. TORFINN/PANOS PICTURES



Feeding future populations means doubling the productivity of neglected but nutritious crops such as yams and green bananas.

## Freeze the footprint of food

**Jason Clay** identifies eight steps that, taken together, could enable farming to feed 10 billion people and keep Earth habitable.

**T**he single largest human impact on our finite planet comes from producing food. By 2050, there will be 2 billion to 3 billion more people on Earth with three times more per capita income, consuming twice as much as now. About 70% will live in cities — more than are alive today. By 2050, we may need three Earths to meet the demands of our consumption. We urgently need to find ways to do more with less.

In the past 18 months, members of non-governmental organizations (NGOs), academia and the private sector have come

together to develop ways to reform the global food system by increasing food production without damaging biodiversity. Groups such as the Global Harvest Initiative ([www.globalharvestinitiative.org](http://www.globalharvestinitiative.org)) and the Sustainable Agriculture Initiative ([www.saiplatform.org](http://www.saiplatform.org)) are working to freeze the footprint of food.

It is a daunting challenge. An estimated 70% of the land that is suitable for growing food is already in use or under some form of protection. For 50 years, farmland has grown at 0.4% a year, at the cost of natural habitat.

In the past decade, as developing economies have grown, this has increased to 0.6% and, with it, more biodiversity has been lost.

Historically, technology has helped to stem this expansion of the agriculture frontier. During the 'green revolution' of the 1960s and '70s, productivity increased at a faster rate than population and consumption, and encroachment was slowed or even halted in many places. Now, technology lags behind rising population and consumption. It needs to catch up, fast.

We will all feel the consequences of an ►

► unhealthy planet, but developing regions will bear the heaviest burden. Nowhere are these realities more pressing than in Africa. The effect of rising food prices has sparked political strife in Tunisia, Algeria and Egypt. Africa is a continent with many complicating factors, and solutions to feeding the planet should be applied there first.

Freezing the footprint of food will require many actors working on several strategies simultaneously. There is no silver bullet. My experiences working with farmers in Latin America, Asia and Africa, and my current role as senior vice-president of market transformation at conservation group WWF, has shown me that we can find common ground with producers big and small to reduce the impact of key commodities.

I have identified eight strategies that, if applied globally and simultaneously, will help to reform the food system and protect the planet. Work has started on each of these 'food wedges', but no group is tackling them all at once. For example, WWF and its partners are directly supporting action on genetics, waste and agricultural carbon. Progress on the others requires more ideas and help, especially in Africa where the challenge is the starkest. Here are some of the goals — and research gaps — as they apply to Africa.

## EIGHT FOOD WEDGES

**Genetics.** Ten crops account for 70–80% of all calories consumed. Only one is on track to double production by 2050. Most estimates suggest that all ten need to double to meet future demand. I'm an environmentalist and am convinced that to increase production, we can't afford to ignore genetics, as long as it is applied in a responsible way. There has been a lot of debate over genetic modification, but there is in fact huge potential in using genetics through traditional plant breeding to select traits — techniques which humans have been using for more than 6,000 years. Now we have twenty-first century technology that allows even faster selection.

In Africa, staple food crops such as yams, plantains and cassava have been relatively neglected by plant breeders. The genomes of these crops should be mapped as a first step towards solutions to doubling or even tripling productivity, and improving drought tolerance, disease resistance and overall nutrient content. Genetic mapping would allow researchers to identify specific traits and markers within a species, and eventually breed plants displaying them. There are plant breeders in Africa prepared to do this.

On 1 July, the African Union formally stated that increasing the productivity of neglected crops in Africa is a priority when it comes to increasing food production there. In June, the African Union's New Partnership for Africa's Development (NEPAD), food company Mars and WWF convened

agriculture experts at a meeting in Washington DC to identify a host of neglected crops in Africa. The group will work with a major scientific institute to sequence the genomes of these crops over the next 3–5 years and then place the information in the public domain. The long-term goal is to train and involve local plant breeders to reduce the time it takes to get acceptable planting materials to farmers.

**Better practices.** For every crop, the best producers globally are 100 times more productive than the worst. Even within nations, producers can be 10 times more efficient than their neighbours, whether they farm maize (corn) in Nebraska or cassava in Nigeria. We will gain most — in terms of food production, increased income and reduced environmental impacts — by improving the poorest-performing producers.

In Africa, it takes too long for better practices to be passed along within the farming community. Traditionally, farmers learn from their parents and from other farmers. Innovation is slow in this closed loop, and governments have scant funding for educating farmers. Through mobile phones, which many African farmers already have, we can help farmers to connect to a shared information hub, allowing one individual to serve many villagers.

Conventionally, such extension systems have been run by governments, but it is not clear if they are up to the task in Africa. Whether provided by the private sector or by government, these systems need to provide value to farmers — increased production, efficiency or net profits. New information hubs must leapfrog, or at least compliment, more traditional extension services.

**Efficiency through technology.** We need to double the efficiency of every agricultural input, including water, fertilizer, pesticides,

energy and infrastructure. It currently takes one litre of water to produce one calorie of food. If we halved the water used and doubled the production we would quadruple the efficiency. The technology exists to do this, and the best producers can already achieve these results.

In Africa, many technologies are two or even three generations behind those used elsewhere. Soil is a great place to start. Increasing organic matter (through root mass and mulching) can rebuild the fertility of soils, double production and halve water usage and other inputs. Measurement is also key — the distance between plants, between rows, the amount (and timing) of fertilizer applied. The biggest challenge may be that many smallholdings in Africa are simply not economical. Fortunately, some technologies are scale-neutral: mulching works even in household gardens.

**Degraded land.** Instead of farming in new areas, we need to rehabilitate degraded, abandoned or underperforming lands. Global goals should be 100 million hectares rehabilitated by 2030 and 250 million by 2050. This means not just halting erosion and degradation but reversing it through the construction of terraces and the planting of trees and grasses. Most farmland in Africa has been degraded over the past century by obsolete practices that were developed when population densities were lower. Ethiopia and South Africa have shown that rehabilitation can work. Each has supported efforts to halt soil erosion, and used a combination of trees, grasses and crops to build up soil organic matter.

**Property rights.** How many farmers will plant a tree or invest in sustainability if they don't own the land, not just for themselves but to pass on to their children? The lack of clear property rights is a significant barrier to food security in Africa, especially in female-led households, which make up the majority of smallholders. By 2020, we should



Can we halve food waste? Greater use of grain silos in Africa could help cut post-harvest losses.

G. OSODI/PANOS PICTURES





The use of farming inputs, including water and fertilizers, needs to be more efficient. This means introducing better tools and practices in places such as Mali.

D. TELEMANS/PANOS PICTURES

aim for 50% of African households to have a title to the lands they cultivate.

Changing this will not be easy, because property rights are controlled by governments. Foreign assistance for economic development should be linked to the establishment of property rights for individuals. The African Union, NEPAD or the World Bank could take the lead in encouraging nations to ensure property rights and to document positive changes on the ground.

**Waste.** Globally, we waste as much as 30–40% of all food produced, or one of every three calories. If we could eliminate waste, we would halve the amount of new food needed by 2050. In rich nations, most food is wasted by individual and institutional consumers.

In Africa, most food waste results from post-harvest losses and lack of infrastructure. One solution could be a one-tonne storage device that safeguards grain and other food, allowing product storage at harvest and until market prices improve. Such a device would need to be collapsible, resealable, locally repairable, and to protect food from moisture, animals, insects and mould. A monetary prize would encourage several prototypes, and a leading institution should champion it. Our goal in Africa should be to cut post-harvest waste in half by 2030.

**Consumption.** One billion people don't have enough food, and yet one billion people eat too much. We need to cut each of these figures in half by 2030, with the most urgent focus on those without enough to eat. About half of these people do not own land or produce their own food; they are split between rural and urban areas, but by 2050 most will live in cities.

About 40% of children under the age of five in sub-Saharan Africa are stunted from malnutrition, and as a result have reduced skills, income and lifespans. The leaves of many common crops in Africa, such as

cassava, sweet potatoes and amaranth are dense in nutrients, but are often not seen as traditional foods and thus not eaten. These leaves should be used to enrich flour in school lunch programmes and, through the education of mothers, in home cooking. The rural poor in Africa have always had access to such wild 'famine foods', but for the urban poor there is no such buffer.

**Carbon.** Soil carbon — or organic matter — is key to conserving farmland for future generations. Indeed, the single best measure of rehabilitated soil is increasing organic matter from less than 0.5% to 2% or more. However, half of the world's top soil, in which most soil carbon resides, has been lost in the past 150 years.

Some analysts suggest that Africa has been losing 1% of soil organic matter every year since the 1960s. This is worse than in any other region of the world, and it results in lower productivity and inefficient use of inputs such as fertilizer and water. Burning (before planting, before harvest or after harvest) decreases soil organic matter. This was an acceptable agricultural practice when land was plentiful and left fallow for many years. But, with rising populations and smaller plots of farmland, practices need to change. Scientists in Australia, for example, suggest that when sugarcane is not burned before harvest, producers save up to 1.5 million litres of irrigation and rain water per hectare, because organic matter retains soil moisture.

Two other approaches would help Africans to conserve their soils. First is a greater emphasis on tree crops and deep-rooted grasses. Trees and grasses build soil carbon and reduce erosion, increasing yields and the efficiency of inputs. Trees can be cash or subsistence crops, and can be assets in their own right (as a source of firewood).

Second, we need carbon markets for agriculture. Retailers or brand-named

companies that purchase sugar, milk, coffee, cocoa or palm oil could also buy the carbon that the farmer sequestered or avoided releasing during production. The carbon would need to be third-party verified and aggregated at a mill or trading house. The goal should be for food producers to sell 1 billion metric tonnes of carbon per year by 2030. This would make food production more sustainable, marginal lands more viable, and producers more financially secure. Over the next year, WWF, with support from the Dutch government and food-linked companies including Unilever, Nutreco and Rabobank, will begin to explore the amount of carbon that could be bundled with commodities and sold in global markets.

### DOUBLE OR BUST

Progress on some food wedges will occur faster than others. But every current system of food production needs to double productivity per hectare. If we cannot double the genetic potential of the 10–15 main calorie crops, on the same amount of land, we will fail to meet rising demand. NGOs and academics do not control the global food system, so instead they must try to change how governments and the private sector think about food production.

Today, most farmers in Africa do not produce enough to feed their own families. No single strategy will solve the global food problem or even ensure sufficient food for Africa. But with the right partnerships, and with improvements across the board, we might be able to feed the world without destroying the planet. ■

**Jason Clay** is senior vice-president, market transformation, WWF, Washington DC 20037, USA.  
e-mail: [jason.clay@wwfus.org](mailto:jason.clay@wwfus.org)





WorldView-2 satellite maps have helped the Tropical Ecology, Assessment and Monitoring Network to operate on a far larger scale.

# Conservation science outside the comfort zone

Researchers like to work on projects that start small and slowly scale up. They must think bigger and faster, says **Sandy J. Andelman**, to tackle today's problems in time.

Six years ago, I leapt from the ivory tower. I left my comfortable job as deputy director of the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara, to do something big and risky: to lead the creation of the Tropical Ecology, Assessment and Monitoring (TEAM) Network, an early warning system for biodiversity loss caused by climate change. The opportunity to work at a global scale with long-term funding prompted my leap. The project was in principle supported for 10 years by US\$43 million from the Gordon and Betty Moore Foundation of Palo Alto, California.

Today, TEAM ([www.teamnetwork.org](http://www.teamnetwork.org)) links 18 tropical monitoring sites in 15 countries in Africa, Asia and Latin America, and continues to grow. At each site, local partners use standardized methods to measure five things: the diversity of trees and woody vines called lianas; carbon stocks; bird and mammal diversity; human-landscape interactions; and climate. All of the data are freely available in near real time. The project is led by Conservation

International (CI) in Arlington, Virginia.

At the beginning, I thought that the best strategy was to set up sites one by one, measuring many things at each site to fully capture local complexity. It soon became clear that adding sites incrementally would not answer our global questions. We needed to sacrifice some local details to get the global system up and running quickly. A new opportunity really brought that message home. In 2009, we gained a boost of funding to monitor ecosystem services. It came with strings attached: there was no time to start small, we needed to start big.

## BIGGER IS BETTER

Initially I fought against getting too big too fast. But I have had a change of heart. I now believe that all conservation scientists need to be thinking and acting more boldly than we are today. If we are to deliver the knowledge we need to manage our hot, crowded, rapidly changing Earth, we need to get outside our comfort zones and take some large, if uncomfortable, steps.

In November 2009, TEAM secured a grant from the Bill and Melinda Gates Foundation in Seattle, Washington, initially for \$435,000, to lead the development and field testing of a set of standard metrics for ecosystem services in areas of agricultural intensification. We had a perfect place for the pilot. TEAM had a well-managed monitoring site in the Udzungwa Mountains National Park in Tanzania. To its south, the Kilombero Valley was targeted by the Tanzanian Ministry of Agriculture and international donors for around \$65 million in investments to double food production over a three-year period. The farmers there depend directly on ecosystem services from the Udzungwa Forest — including water, wood for fuel, bush meat and protection from erosion — for their agricultural production and livelihoods.

For the pilot, we planned to monitor the following in both the Udzungwa and the Kilombero Valley: biophysical properties (from biodiversity to water quality and climate); agricultural productivity (for example, areas planted and crop yield); livelihood measures



(such as household income and under-five mortality rates); and resilience of natural and human systems to climate variability.

Many ecological projects tackle only a few hectares. TEAM sites average about 3,800 square kilometres, with standardized field measurements covering around 400 km<sup>2</sup>. The pilot area was a daunting 5,000 km<sup>2</sup>.

Two months after receiving the grant, Prabhu Pingali, deputy director of agricultural development at the Gates Foundation, told me that our envisioned pilot project was far too small and slow. The foundation needed a pilot that, within a year, would cover much of southern Tanzania — a seemingly impossible area of about 335,000 km<sup>2</sup>.

### WEIGHING THE TRADE-OFFS

Pingali's rationale was compelling. There have never been as many hungry people on the planet as there are today, most of them smallholder farmers in developing countries. Conservation scientists need to provide good data and methods for monitoring change, at relevant scales, as soon as possible. Otherwise, decisions about agricultural development will continue to be made without properly weighing the trade-offs and synergies between agriculture, nature and human livelihoods. The Alliance for a Green Revolution in Africa (AGRA), a Gates Foundation partner, aims to double food production in three years in southern Tanzania and in regions of Mozambique, Ghana and Mali. It needed good baseline data and monitoring techniques for ecosystem services immediately — not in a few years. Pingali told me: we can give you more resources, but we can't give you more time.

This was a troubling challenge. I was comfortable with the model of: start small because resources are scarce; carefully test methods and sampling design; publish initial results; and then iterate and scale upwards slowly and carefully, with peer review informing every step. I explained to Pingali that 'we' — Conservation International, the TEAM Network, the conservation-science community and I — had no idea how to monitor ecosystem services consistently at such a huge scale. But I took another leap and said we would try.

Eighteen months later, we have completed the pilot project. Thinking about large-scale methods from the outset pushed us towards practical, innovative technologies and partnerships.

For example, we used high-resolution imagery from the WorldView-2 satellite to assess fine-scale land-cover patterns, from tilled land to forest. These kinds of data are

expensive, but they are needed for large-scale work. We distributed georeferencing camera phones to local people and researchers. We used their photos to validate and supplement remote-sensing images. We partnered with the Tanzanian National Bureau of Statistics and the World Bank (which together run a gold-standard annual survey of livelihood and agricultural management) to integrate their social data with our biophysical measures — of water availability, for example. And we adopted a protocol for measuring soil organic-carbon levels from the African Soil Information Service and the World Agroforestry Centre in Nairobi, rather than reinventing the wheel. We are now replicating this strategy in Rwanda and plans are under way to cover the rest of sub-Saharan Africa, Asia and Latin America.

There have been bumps in the road. We still don't have the right algorithms to automate processing of the high-resolution, remote-sensing images, for example. It takes one highly skilled analyst two weeks

pace of the peer-review process and the outdated reward systems of our institutions. But these are obstacles that we must overcome.

Detractors may argue that large-scale approaches by definition can't capture the level of detail needed to fully understand complex systems. But development decisions are large and often can't handle that level of complexity. The key is to select the measures

that are relevant to both science and policy.

I am not saying that conservation researchers should give up scientific and analytical rigour.

But we do need to trade

in our slow, incremental models of funding and investigation for something bolder. In 2010, the entire Division of Environmental Biology at the US National Science Foundation gave out 702 grants, averaging just over \$212,000 each, with an average duration of two years. Forty-nine per cent of these went to lone investigators. Planning for the US

National Ecological Observatory Network (NEON) began 11 years ago, and has cost more than \$90 million to date. But this comes from a pot of money devoted to construction of large facilities, not to ongoing research, and it has not yet moved to implementation.

The NSF funding model does not support global-scale conservation science for a rapidly changing world. We need to look to non-traditional funding sources such as the private sector, and actively work to set up consortia of donors.

What will it cost to scale up? One 2008 estimate suggested that a global monitoring network for biodiversity and ecosystem services would cost \$309 million to \$772 million a year (R. J. Scholes *et al. Science* **321**, 1044–1045; 2008). On the basis of my experience, I believe that we can

create a scientifically credible, policy-relevant global network for more like \$10 million a year, by integrating proven methods from successful networks — such as TEAM, NEON and the Digital Soil Map of the World — with the full arsenal of innovative informatics tools and mobile technologies. Such a network will not measure everything, everywhere, but it should be able to provide the targeted data that policy-makers need. It is time for conservation scientists and donors to step up to this challenge. ■

**Sandy Andelman** is vice-president and director of the Tropical Ecology, Assessment and Monitoring (TEAM) Network, with Conservation International, Arlington, Virginia 22202, USA.  
e-mail: s.andelman@conservation.org



A pilot project in Tanzania using highly targeted on-the-ground sampling.

to process an image covering 100 km<sup>2</sup>. But on the whole it worked. We established, very quickly, a baseline measure of the system before agricultural intensification.

### BIG AMBITIONS

Most conservation science today isn't ambitious enough. We are informing battles, but we are not providing the knowledge needed, at the scale needed, to win the war. For example, global policy-makers and national governments are trying to produce robust estimates of forest carbon stocks to assist in managing emissions and carbon sequestration. But the error in regional- and global-scale estimates of forest carbon is as high as 50%, mainly because of limitations in the scale of measurements. There are some obvious constraints: limited funding, the slow





Red lines on an 1832 map show how cholera spread from India to Europe and North America along major trade routes.

## EPIDEMIOLOGY

# Charting the spread of sickness

A history of disease mapping shows that despite technological developments, little has changed in 500 years, finds **Andy Tatem**.

“Ill-clothed, ill-fed, uncleanly in their persons and dwellings, subsisting chiefly on indigestible and unwholesome foods and in the habit of using pernicious drinks.” So wrote physician John Hamett in the *London Medical Gazette* — describing not modern teenagers, but rather the people within nineteenth-century society who were thought by many to cause disease. ‘Putrid effluvia’, ‘foul personal habits’ and ‘noxious miasmas’ were held liable for sickness until the late 1800s. And from the seventeenth century, maps of outbreaks were used as tools in debates — with clusters of cases in the poorest, smelliest or lowest parts of town apparently confirming suspicions that the disease in question was caused by poor people, sewage or altitude.

In *Disease Maps*, medical geographer Tom Koch tours the history of disease mapping, focusing on plague, yellow fever, cancer and cholera. He explains how maps of each have both increased our understanding of epidemiology and fuelled now-discredited theories. Mapping techniques have advanced hugely — today’s surveys are located using

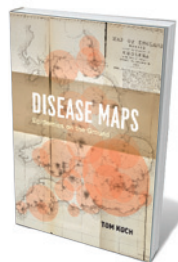
the Global Positioning System and analysed using geographic information systems. Yet Koch’s main message is that the politics, use and interpretation of spatial disease data have changed little in the past 500 years.

In a world where we can download a mobile-phone application that locates for us the nearest outbreak of Rocky Mountain spotted fever, it is easy to forget the poor understanding of pathogens we once had. In the late seventeenth century, a good 200 years before bacteria and viruses were properly understood, it was thought that there were many ‘plagues’ — a sickness that appeared only in warm seasons; another that was caused by noxious smells from slums; and others that moved from city to city with

ships and travellers. Disease maps were drawn up but generally proved inconclusive.

Such uncertainty lingers on in scientific debates over the socioeconomic drivers of epidemic spread, or the effects of climate change on vector-borne diseases such as malaria. Disease mapping has allowed us to detect potential sociological or environmental reasons for outbreaks through their spatial association. However, maps are easily misinterpreted because of the challenge of untangling correlation and causation. The book is strewn with examples of outbreaks that have been found to cluster with a particular factor, leading to claims of an ultimately unfounded causal link. Such was the case for cholera, until one of the most famous disease-mapping studies was undertaken.

John Snow’s plotting of the 1854 Broad Street cholera outbreak in London suggested the waterborne nature of the disease. His halting of the epidemic by removing the handle of the water



**Disease Maps: Epidemics On the Ground**  
TOM KOCH  
University of Chicago  
Press: 2011. 368 pp.  
\$45, £29

**NATURE.COM**  
Old Islamic maps of science:  
[go.nature.com/qdzj1h](http://go.nature.com/qdzj1h)



pump around which the cases clustered is a popular tale, but Koch reveals that the conventional version is oversold. Snow's idea on the waterborne transmission of cholera was later proved correct, but his work, and that of many other mappers of the time, merely revealed the conditions in which cholera thrived, not that it is caused by a bacterium, *Vibrio cholerae*.

Other cholera case studies in the book include an 1831 continental-scale meta-analysis of the outbreaks that spread across Europe in the early nineteenth century — a forerunner of modern large-scale mapping projects. Koch highlights colonial Indian maps of cholera spread that were of limited use because 'natives' were not included, and London neighbourhood mapping battles that aimed to prove all kinds of transmission mechanisms. Each contains lessons for today.

*Disease Maps* is well researched, and packed with beautifully reproduced epidemiological maps and colourful tales of the arguments and insights each one sparked. Yet it skirts recent advances in our knowledge of disease spread. There is no mention of the

state-of-the-art tools that could have settled some of the historical debates.

**"According to London's 1667 bill of mortality, 11 people were killed by itches and one died from fainting in the bath."**

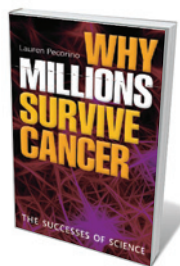
Developments include new phylogeographic techniques for reconstructing spatial and temporal evolutionary histories of the spread of pathogens such as influenza, and Bayesian statistical models for generating uncertainty

maps to accompany those charting endemic diseases such as malaria. Mobile-phone data are also providing unprecedented opportunities for exploring short-term spatial epidemiological dynamics by tracking human travel patterns.

At a time when resources are flowing into the fields of health metrics and epidemiological mapping, *Disease Maps* shows that some things never change. The epidemic drivers of increased trade, rapid urbanization, inequalities and civil unrest are as strong today as they were during the seventeenth-century outbreaks of plague. Koch's book takes us back to the dawn of disease statistics and mapping, when — according to London's 1667 bill of mortality — 11 people were killed by itches, one person died from fainting in the bath and many more perished when their stomachs simply 'stopped'. ■

**Andy Tatem** is an assistant professor at the Emerging Pathogens Institute and in the Department of Geography of the University of Florida, Gainesville 32610, USA.  
e-mail: andy.tatem@gmail.com

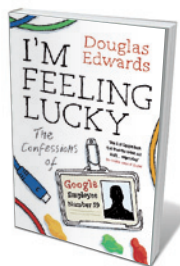
## Books in brief



### Why Millions Survive Cancer: The Successes of Science

Lauren Pecorino OXFORD UNIVERSITY PRESS 256 pp. £16.99 (2011)

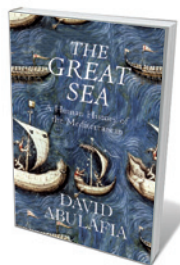
One in three of us can expect to have cancer during our lifetimes. But the prognosis is good, according to molecular biologist Lauren Pecorino. More people are surviving as better treatments come on line, thanks to advances in science and medicine. Relating the latest scientific evidence, she details for the general reader how models of cancer and knowledge of how the body defends itself against tumours have improved, and shows how the disease is better managed today. The book also examines the science that lies behind various lifestyle factors that contribute to cancer risks.



### I'm Feeling Lucky: The Confessions of Google Employee Number 59

Douglas Edwards ALLEN LANE 432 pp. £20 (2011)

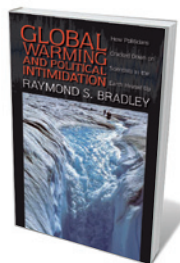
Douglas Edwards was Google's first director of marketing and brand management — employee number 59 — a post he held from 1999 to 2005. In his book he offers a peek inside the Googleplex, giving an intimate portrayal of the innovative company's unique culture and how it developed. He describes how the firm's founding duo, Larry Page and Sergey Brin, have encouraged a non-hierarchical management structure and fostered a creative ethos. He also gives sage management advice, explaining, for example, why you should always hire someone smarter than yourself.



### The Great Sea: A Human History of the Mediterranean

David Abulafia ALLEN LANE 816 pp. £30 (2011)

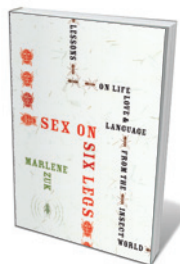
The Mediterranean Sea has witnessed the meeting of many civilizations throughout history. In this magnum opus, historian David Abulafia tells the tales of the diverse peoples that have lived around the Great Sea, portraying their trade and battles and emphasizing their varied languages and societies. From the Trojan Wars in the eleventh or twelfth centuries BC to the Grand Tours of the nineteenth century AD, and from the history of piracy to the spread of religions and modern tourism, he paints the Mediterranean as an epicentre of human history.



### Global Warming and Political Intimidation: How Politicians Cracked Down on Scientists As the Earth Heated Up

Raymond S. Bradley UNIVERSITY OF MASSACHUSETTS PRESS 168 pp. \$19.95 (2011)

In 2005, US-based climate scientist Raymond Bradley found himself in the middle of a political maelstrom. Sceptical congressmen demanded that he and his co-authors, who had published the famous 'hockey stick' graph of rising atmospheric carbon dioxide levels, hand over their data and declare their funding sources. Bradley relates this troubling episode and expresses his concern that some politicians are seeking to suppress climate science.



### Sex on Six Legs: Lessons on Life, Love and Language from the Insect World

Marlene Zuk HOUGHTON MIFFLIN HARCOURT 272 pp. \$25 (2011)

Most people recoil from creepy crawlies; biologist Marlene Zuk explains their scientific allure in her latest book. She relates that insects are more numerous than any other type of animal, accounting for 80% of species. And she describes how flies, ants, wasps and their ilk mate, care for their offspring, hunt and defend themselves. Her entertaining, no-nonsense prose is packed with colourful examples.

## TECHNOLOGY

# The spacesuit unpicked

Margaret Weitekamp reflects on how fashion influenced astronautical attire for the Apollo missions.

When Neil Armstrong took “one small step” on to the Moon’s surface in 1969, his soft, white, layered spacesuit insulated, cooled, pressurized and protected his body. But it was not what many people had thought he would be wearing. Most engineers imagined lunar landing suits as hard, man-shaped, jointed shells containing pressurized environments — essentially, individual ambulatory spacecraft. The iconic suit, known as the A7L and designed by the International Latex Corporation (ILC) then in Dover, Delaware, accomplished the same task with carefully stitched layers of innovative materials. Synthetic, rubberized or reinforced, each was chosen to impart flexibility, strength, insulation or protection without excessive weight or bulk.

Using the 21 layers of the A7L as his inspiration, in the same number of chapters, architect Nicholas De Monchaux considers the social, cultural and political contexts of the iconic suit. He sees the hand-crafted garment as an essential counterpoint to today’s prevailing emphasis on systems, so common

**NATURE.COM**

For more on conserving NASA’s spacesuits:  
[go.nature.com/kdsdm1](http://go.nature.com/kdsdm1)

in engineering, architecture and design. Systems thinking considers humans as one factor in a broad engineering schema rather than dealing with humanity’s complexity and variability, de Monchaux argues. He contends that such abstract thinking is not adaptable enough to master the realities of human life, at any scale. Instead, he suggests that the Apollo spacesuit offers a model for creating complex, responsive designs for other human environments, such as cities on Earth — “our only enduring spaceship”.

Unlike many other books on the space race, *Spacesuit* connects the technical story with the broader history of the period, linking fashion with the military-industrial complex. De Monchaux includes the expected histories of high-altitude ballooning and its perils (including hypoxia and the bends), the development of partial- and



**Spacesuit: Fashioning Apollo**  
NICHOLAS DE MONCHAUX  
MIT Press: 2011.  
250 pp. \$34.95

full-pressure suits and the experiments of early aeromedical researchers. But he also weaves in Christian Dior’s New Look.

Dior’s extravagant dresses of 1947 electrified the postwar world with their full skirts that, in contrast to those made during wartime fabric rationing, used many yards of fabric. ‘New Look’ became shorthand for describing sweeping changes in fields as diverse as plastics, criminal justice, roads and politics. De Monchaux unites these diverse topics through his analysis of the ‘new look’ in defence planning, a concept that drove changes in postwar agencies such as NASA.

It is at the core of the story, however, that *Spacesuit* gets most interesting — and most controversial. Those who know the history of NASA’s spacesuit contracts will bristle at de Monchaux’s conflation of the ILC and Playtex, maker of bras and girdles. In 1947, the materials company ILC split into four divisions. One was Playtex. But it was another division that, in 1962, became a subcontractor to Connecticut-based Hamilton Standard on a project to create a lunar space suit.

That the book pays little attention to other aerospace corporations working on spacesuits, including Hamilton Standard and the David Clark Company of Worcester, Massachusetts, will also irk some readers. With much personal and corporate pride still tied up with the production of these famous suits, *Spacesuit* will be seen as taking sides. But de Monchaux’s book is not intended to be the authoritative history of spacesuit contracts. For that, see Kenneth Thomas and Harold McMann’s *US Spacesuits* (Praxis, 2005), which excels in details but lacks readability.

De Monchaux has an ear for a good story and affection for the historical characters. In 1967, after the *Apollo 1* cabin fire that killed three astronauts, NASA revisited its spacesuit contracts. In response to the new call for proposals, the ILC’s soft suit design won a fierce competition to be the Apollo programme’s Moon-walking suit. De Monchaux frames it as a David versus Goliath story, in which the “hard-knocks” engineers of the upstart ILC dug deep to outperform the aerospace behemoths. Through trials that included playing American football in the suit, the ILC garment’s flexibility, compactness and durability won the day. *Spacesuit* offers a broad and creative appraisal of that suit’s many contexts, encouraging readers to consider technology as design, shaped by the circumstances of its time, unfailingly and elegantly layered and crafted to serve a purpose. ■

**Margaret A. Weitekamp** is curator of the social and cultural dimensions of spaceflight at the National Air and Space Museum, Smithsonian Institution, Washington DC, USA.  
e-mail: [weitekampm@si.edu](mailto:weitekampm@si.edu)



Buzz Aldrin (left, just seen), Michael Collins and Neil Armstrong wore hand-stitched A7L suits for Apollo 11.

NASA/KENNEDY SPACE CENTER



CAROLYN KASTER/AP



## Q&A Richard Berendzen

# The sci-fi adviser

Richard Berendzen is director of NASA's Space Grant Consortium in Washington DC, and advised on the science-fiction film *Another Earth*, winner of the Alfred P. Sloan Feature Film Prize for science at this year's Sundance Film Festival. On the film's North American release, he talks to Nature about parallel worlds and the future of human space exploration.

### How did you get involved with *Another Earth*?

Director Mike Cahill and co-writer and star Brit Marling approached me after they had listened to *Pulp Physics*, a set of audio tapes I'd made in 2001 about the history of astronomy. They didn't have a script for their film at the time, so they asked me some scientific and philosophical questions. They recorded my responses and later used my voice as the narrator. To have created such a thought-provoking film with these limited resources, they almost seem to be from another Earth. Cahill's economy with the script, dialogue and editing produces a haunting effect. And Marling's face projects a range of emotions without uttering a word.

### What is the plot?

A duplicate Earth is discovered in our Solar System. Marling plays an astrophysics student who is distracted by the new planet as she drives home. She crashes, killing a composer's wife and children. She applies to visit the sister planet, where her mirror-self presumably avoided the accident. The film raises questions about the human condition, such as how do you apologize for the unforgivable? How long should a person do penance for a dreadful act? What if you could meet yourself?

### What is the science behind the film?

The physics of string theory can lead to quantum-mechanical models in which parallel universes arise. There could be one or an infinite number of them. They might be only a millimetre away from us. And some of them could, in theory, contain another Earth and another you. The nearest potentially viable planet we have found is Gliese 581e, which is about 6 parsecs away from us. That great distance prohibits travel, so in the film the second planet is portrayed as close. One of the film's strengths is how it prompts debate about diverse facets of science.



### *Another Earth*

DIRECTED BY MIKE CAHILL

Released on 22 July in North America.

### How did you come to study astronomy?

As a boy I looked at the stars and wondered what they were. Science-fiction films of the 1950s such as *Destination Moon* and *The Day the Earth Stood Still* had a strong impact; they inspired me to find out more. One transformative night while I was studying physics at

the Massachusetts Institute of Technology, I went to the Harvard Observatory and saw the first photograph ever taken of the Moon with a telescope. It was primitive by modern standards, but back then it looked dazzling. I found astronomy more interesting than physics because you could study everything from quantum physics to relativity on the grand scale. The history of astronomy is interwoven with the history of human consciousness.

### What lies ahead for NASA?

NASA's future is not clear to me. When Sputnik was launched, the United States was shaken. NASA was formed overnight, and we unmistakably won the space race. We sent out small craft to take close-up photographs of other planets. We sent out robotic landers. But you run out of Solar System after a while. NASA became a victim of its own success. Take the International Space Station: what do you do with it once it is built? You can test human health under weightless conditions, and you can use it as a launching pad to reach the Moon or Mars. But who wants to spend hundreds of millions of dollars sending people to Mars when high unemployment in the United States means you can't get a job?

### Does human space exploration have a future?

There's a drama and a romance to human space flight. But is it worth it? Robotic missions are cheaper and safer. They produce good scientific data and images that the public finds inspiring. If we're going to explore space using humans, we have to learn to live off the Universe. It is hugely expensive to ship water to the Moon. But NASA probes have detected water at the Moon's poles, and we think there is enough slush there to sustain a full exploratory crew for decades. If you've got water, you can break it apart to use the hydrogen as rocket fuel and the oxygen to breathe. If you had a nuclear reactor to burn helium-3, you could have free electrical energy. In principle, you could even leave our Solar System using a ram jet that sucks in interstellar dust as fuel.

### Do younger people take space for granted?

When I was young, space was new and everything was possible. Nothing surprises today's youngsters. They've grown up with so much technology that it takes a great deal to get a 'gee whiz' out of them. But when I start raising questions about life on other planets, there is silence in the lecture hall. Astronomy can teach awe and humility. After Isaac Newton wrote the *Principia*, he was asked: "What is gravity?" He replied: "I frame no hypotheses" — which means, 'beats me'. You study the cosmos your whole life, and then you realize, to paraphrase Newton, I'm like a child at the seashore with the vastness of the ocean of truth around me. ■

INTERVIEW BY JASCHA HOFFMAN

FOX SEARCHLIGHT PICTURES

# CORRESPONDENCE

## End invasive chimp research now

We disagree with your framing of chimpanzee research around past 'handsome' benefits to humankind (*Nature* **474**, 252; 2011). These do not justify invasive experimentation in the future. The current pace of advancements in biomedical research technology means that the time has come to end invasive research on chimpanzees.

Biomedical research is replete with examples in which reliance on relatively crude, time-consuming and expensive animal models has given way to quicker and more precise non-animal methods. For example, drug maker Eli Lilly struggled to find enough rabbits to test insulin for potency and safety when it was first purified; those tests soon gave way to mouse studies and then physicochemical methods.

Such reliance on animals was not always scientifically necessary. Take the Sabin live polio vaccine: tests in chimpanzees were simply part of a need to reassure nervous regulators that it was safe. The crucial development was cultivation of the virus in human cells — a discovery that received a Nobel prize.

The past decade has seen dozens of reports of non-animal techniques that explore the biology and pathology of hepatitis C virus (HCV), including a new *in vitro* technology that allows replication of the virus from infected patients (M. Buck *PLoS ONE* **3**, e2660; 2008). GlaxoSmithKline runs an HCV programme that does not use chimpanzees (see [go.nature.com/eyp3wk](http://go.nature.com/eyp3wk)).

Some argue that chimpanzees are still needed for testing monoclonal antibody therapeutics, but this does not stand up to scrutiny. Of the 35 monoclonal antibody therapeutics approved so far by the US Food and Drug

Administration, only three involved chimpanzee testing, and two of those were withdrawn because of side effects or lack of effectiveness (R. H. Bettauer *ALTEX* **28**, 103–116; 2011).

**Andrew Rowan, Kathleen Conlee, Raija Bettauer** *The Humane Society of the United States, Washington DC, USA.*  
[arowan@humanesociety.org](mailto:arowan@humanesociety.org)

## Indian vaccine study clarified

As director of the human papillomavirus (HPV) vaccines project at the global health non-profit organization PATH, I wish to clarify some important points relating to your News story on the inquiry committee's investigation of the HPV vaccine study following the deaths of participants (*Nature* **474**, 427–428; 2011). There were seven deaths (among nearly 24,000 girls), not four, and the inquiry committee found that none of the deaths was related to the vaccine (five definitively, two considered unlikely).

Contrary to your headline indicating that the ethics of the study have been criticized, the inquiry committee's report concludes: "There has been no major violation of any ethical norm in the conduct of the study."

PATH believes that the HPV vaccine study under review was not a 'clinical' trial because no clinical outcomes were measured. The product had already undergone clinical trials in India and elsewhere, and had been licensed and made available in the private sector throughout the country. Even so, safeguards such as ethical review and informed consent were built into the study.

You quote Jacob Puliyel's opinion that not enough is known about the burden of HPV-related cervical cancer in India. However, the World Health

Organization estimates that India shoulders more than one-quarter of the global burden of cervical-cancer mortality (about 73,000 of 275,000 deaths per year). The HPV types countered by the vaccines in question account for the majority of these cases. It was in part owing to the high prevalence of cervical cancer in the country that PATH included India among its four study sites (similar studies were conducted in Peru, Uganda and Vietnam).

Prevention methods such as HPV vaccination and screening alternatives (by visual inspection or HPV DNA testing) could substantially reduce the mortality in middle- and low-income countries, which suffer about 88% of the burden of cervical cancer, to the low levels now common elsewhere.

To increase cost-effectiveness, we need evidence of the best way to roll out these interventions. Studies such as those completed in India (see [www.rho.org](http://www.rho.org)) will enable the acquisition of such data.

**Vivien Davis Tsu** *PATH, Seattle, Washington, USA.*  
[vtsu@path.org](mailto:vtsu@path.org)

## Café science for kids

A science café, Zabuki, that we launched for children in the Dutch town of Deventer in 2008 is hugely successful — attracting around 70 schoolchildren every month ([www.zabuki.nl](http://www.zabuki.nl)). We also organize an annual two-day science festival with the local teacher-training schools, to which some 800 children came last year. We urge other towns to launch similar initiatives to encourage more kids to engage with science and technology.

Like science cafés for adults (*Nature* **399**, 120; 1999), Zabuki is run by volunteers — usually parents. Trainee primary-school teachers also help regularly as part of their assignments. The

children (aged 7–12) suggest themes that are developed in smaller workshops with the help of a local expert. Local companies volunteer materials and expertise.

The cafés offer children a hands-on scientific challenge in a stimulating, non-school context, where they can explore themes voluntarily that they might not otherwise investigate or discuss. Topics have included robotics, architecture and life in a ditch.

Children pay €4 (about US\$6) towards expenses and refreshments; the remainder of the €700 cost per café is sponsored by the region. Volunteers spend 2–8 hours a month organizing the café.

**Anne M. Dijkstra** *University of Twente, Enschede, the Netherlands.*

[a.m.dijkstra@utwente.nl](mailto:a.m.dijkstra@utwente.nl)

**Henk Van Voorthuizen** *Abilene, nl, Deventer, the Netherlands.*

**Mark Van Zijtveld** *Deventer, the Netherlands.*

## Rise in scientists returning to China

Jun Li blames China's rigid citizenship regulations for hindering the return of Chinese scientists from abroad (*Nature* **474**, 285; 2011). This is an oversimplification.

Chinese scientists and engineers are much in demand abroad. But because of China's economic development and the downturn in the West, the number of returnees has risen sharply (see [go.nature.com/dfiljeo](http://go.nature.com/dfiljeo)). Taiwan also had a surge of trained scientists returning home during its economic boom in the late 1980s and 1990s.

Given China's buoyant economy, relaxation of China's citizenship regulations is unlikely to happen soon.

**Xiaoming Li** *Tennessee State University, Nashville, Tennessee, USA.*  
[xli1@tntstate.edu](mailto:xli1@tntstate.edu)



## Spin flips with a single proton

**For the first time, spin flips of a single trapped proton in free space have been observed. This is a major step towards a million-fold improved test of matter–antimatter symmetry using a nuclear magnetic moment.**

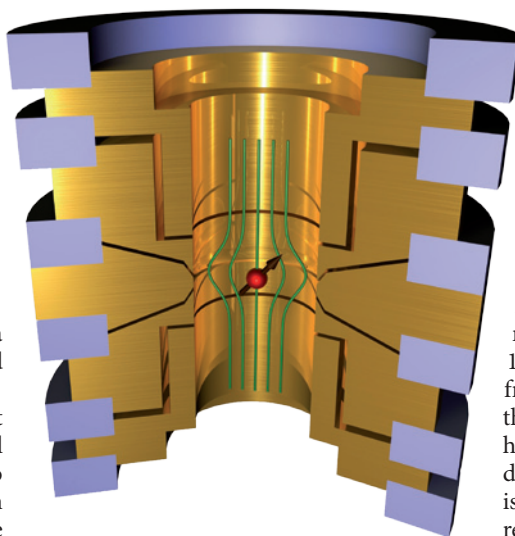
RAINER BLATT

The world, as we know and feel it every day, consists of matter. Whereas the Big Bang is considered to have created both matter and antimatter in equal quantities, the present-day Universe clearly seems to display a huge asymmetry: antimatter is rarely observed, and if it is, it's only in highly exotic environments or in some radioactive reactions. Writing in *Physical Review Letters*, Ulmer *et al.*<sup>1</sup> describe an experiment that paves the way for a high-precision test of the theoretically expected matter–antimatter symmetry.

Matter–antimatter symmetry is the most fundamental symmetry in the standard model of elementary particle physics. According to this symmetry, under a CPT transformation — which is a simultaneous inversion of the particle properties charge (C) and parity (P), and a reversal of time (T) — an antiparticle behaves exactly like its mirror-image particle. Ultra-high-precision tests of the CPT symmetry have been performed in different physical systems (for example, with mesons, leptons and baryons) by comparing the properties of particles and their antiparticles. Yet, until now a violation of this symmetry has never been observed.

One of the most fundamental precision tests of CPT — the comparison of the magnetic moment of a single proton and that of its antiparticle, the antiproton — has yet to be performed. Such a test is extremely challenging because it requires ultra-high-precision spectroscopy of, at best, single and unperturbed protons and antiprotons. And this is where Ulmer and colleagues' study comes in. Their experiment allowed them to flip the spin of a single proton, and this will enable a precision measurement of the proton's magnetic moment to be made. What's more, the experimental set-up can be readily applied to antiprotons, and will eventually provide a precision CPT test.

To observe the spin flips, Ulmer *et al.*<sup>1</sup> stored a single proton in a Penning electromagnetic trap at cryogenic temperatures. The trapped proton oscillates with three main frequencies, one of which — the axial frequency, which is associated with the proton's oscillation along the direction of the Penning trap's magnetic



**Figure 1 | Penning trap used by Ulmer *et al.*<sup>1</sup> to measure proton spin flips.** A strong magnetic-field inhomogeneity (green) couples the proton's spin direction (arrow) to the frequency with which the proton (red) oscillates along the direction of the trap's magnetic field. This frequency is measured to monitor the single-proton spin flips, which are induced by an external radiofrequency field (not shown). Gold electrodes are held in place with insulating sapphire structures (blue).

field — is actually measured in the experiment. To obtain a dependence of the axial frequency on the proton's magnetic moment, the authors added a magnetic-field inhomogeneity to the otherwise homogenous magnetic field of the Penning trap (Fig. 1). In this 'magnetic bottle', the spin direction of the proton shifts the axial frequency, which can be subsequently measured with high sensitivity by means of a superconducting detection system.

The spin-dependent frequency shift observable in the proton's axial oscillation in the (inhomogeneous) Penning trap is due to the 'continuous Stern–Gerlach effect', which was introduced by Nobel laureate Hans Georg Dehmelt<sup>2</sup>. The shift allows the detection of the spin direction of a single trapped charged particle. This method was previously used<sup>3–6</sup> with great success for precision measurements of the magnetic moment of the electron, of the positron and of bound electrons. Now, for the first time, the technique has been successfully applied to a

single proton, whose magnetic moment is almost a thousand-fold smaller than that of the electron. The proton's extremely small magnetic moment had defied its measurement with this method for several decades.

Trapping a single proton in a cryogenic Penning trap allows the particle to be stored for months, and is ideally suited for a high-precision experiment. In the authors' experiment<sup>1</sup>, the large magnetic-field inhomogeneity changed the homogeneous magnetic field by 1 tesla within a distance of about 1 millimetre, and a spin flip changed the axial frequency of the single trapped proton by less than one part in a million — that is, 200 millihertz out of 680 kilohertz. The demanding detection of this tiny frequency shift, which is the signature of the proton's spin direction, represents a real experimental feat.

To drive the spin flips, Ulmer *et al.* used the magnetic component of a radiofrequency signal. If the signal is resonant with the energy difference between the two orientations of the magnetic moment, the spin-flip probability has a maximum of 50%; it decreases if the driving-signal frequency is detuned off-resonant. By measuring the spin-flip probability as a function of the driving-signal frequency, the authors were able to derive the value of the proton magnetic moment. To achieve better measurement precision, they will use a high-precision section of their trap arrangement together with methods that have already been tested<sup>7</sup>. This will allow for a million-fold improved test of the matter–antimatter symmetry using a trapped (anti)proton.

In addition to the exciting prospect of a new high-precision test of the matter–antimatter symmetry, the method can also be applied to directly measure magnetic moments of light atomic nuclei. Together with spectroscopic data, such measurements will contribute to a deeper understanding of nuclear size effects (on atomic spectra) and of the distribution of magnetic moments in atomic nuclei. ■

**Rainer Blatt** is at the *Institut für Experimentalphysik, Universität Innsbruck, and the Institut für Quantenoptik und Quanteninformation, Austrian Academy of Sciences, A-6020 Innsbruck, Austria.*  
e-mail: rainer.blatt@uibk.ac.at

1. Ulmer, S. *et al.* *Phys. Rev. Lett.* **106**, 253001 (2011).
2. Dehmelt, H. *Proc. Natl Acad. Sci. USA* **83**, 2291–2294 (1985).
3. Van Dyck, Jr. *et al.* *Phys. Rev. Lett.* **59**, 26–29 (1987).
4. Hermanspahn, N. *et al.* *Phys. Rev. Lett.* **84**, 427–430 (2000).
5. Verdu, J. *et al.* *Phys. Rev. Lett.* **92**, 093002 (2004).
6. Hanneke, D., Fogwell, S. & Gabrielse, G. *Phys. Rev. Lett.* **100**, 120801 (2008).
7. Häffner, H. *et al.* *Phys. Rev. Lett.* **85**, 5308–5311 (2000).

## NEUROSCIENCE

# Imprinting in the brain

**A gene is considered to be imprinted if only the copy inherited from the mother or from the father is expressed throughout life. But one imprinted gene, *Dlk1*, disobeys this rule during postnatal neurodevelopment. SEE LETTER P.381**

EDWIN C. OH & NICHOLAS KATSANIS

In the adult mammalian brain, neural stem cells continually give rise to new neurons. There is hope, therefore, that these cells might also be of use for treating brain injuries and neurological disorders. Although the molecular signals that establish and maintain neural stem cells are starting to be understood<sup>1,2</sup>, the roles of genetic determinants and environmental cues in these processes remain largely mysterious. On page 381 of this issue, Ferrón and colleagues<sup>3</sup> show how one gene, *Dlk1*, is necessary for regulating the numbers of mouse neural stem cells soon after birth and in adulthood.

Multicellular organisms inherit equivalent complements of maternal and paternal chromosomes. Nonetheless, the expression of genes from either the maternal or the paternal copy of a chromosome can be regulated by DNA methylation and post-translational modification of the DNA-associated histone proteins in a heritable process called genomic imprinting<sup>4</sup>. In the brain, more than 1,300 transcripts are imprinted<sup>5–7</sup> — a feature that is remarkably distinct from other tissue-specific expression profiles and indicates the brain's particular sensitivity to imprinting. *Dlk1*, which also has roles in the regulation of fat tissue, blood cells and skeletal-muscle development, is one such imprinted gene in the mouse brain, as only its paternally inherited copy is expressed<sup>8</sup>.

Given the significance of imprinted genes in the nervous system, Ferrón *et al.*<sup>3</sup> investigated the role of *Dlk1* in embryonic, postnatal and adult mouse neurogenesis in the subventricular zone (SVZ) of the lateral ventricles. The SVZ is one of two discrete regions in the adult brain where neurogenesis occurs, the other being the subgranular zone in the dentate gyrus of the hippocampus — a structure implicated in learning and memory.

In neurogenic areas, the authors detected *Dlk1* expression both in mouse embryos and after birth, with levels peaking at postnatal day seven (P7) — a transition period between the development of the embryonic and the mature SVZ. However, whereas mice lacking *Dlk1*

(*Dlk1*<sup>−/−</sup> mice) showed no major differences in the number of proliferating cells in their brains' germinal regions at embryonic stages, at P7 they had more proliferating neural stem cells (NSCs) compared with normal mice. Notably, the authors also observed more proliferating NSCs at P7 in mice carrying a non-functional *Dlk1* copy of either maternal or paternal origin — a finding that suggests that imprinting is irrelevant in postnatal neurogenesis.

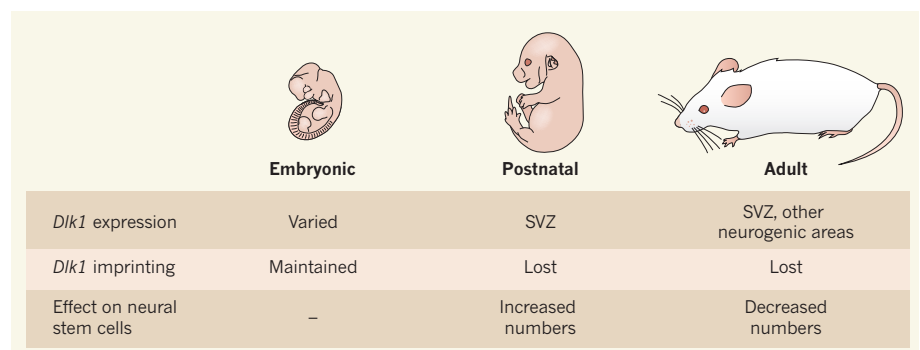
In the adult brain, loss of *Dlk1* was also associated with a precipitous decline in the number of NSCs and fewer neurons migrating from the SVZ towards their final destination (the olfactory bulb). These results indicate that disruption of NSC quiescence at early postnatal stages leads to their depletion — and so to depletion of neurons arising from them — in the adult brain.

When stem cells harvested from the brain are maintained in tissue culture, they form spherical clusters called neurospheres<sup>9</sup>. To examine the mechanism by which *Dlk1* might affect NSC maintenance, Ferrón *et al.* examined whether the neurosphere yield is different between normal and *Dlk1*<sup>−/−</sup> mice. Consistent with their *in vivo* data, they observed a transient increase, followed by a decline, in the number of primary neurospheres from the mutant SVZ tissue over time. This result supports the hypothesis that a neurogenic continuum is created at an early postnatal period and is maintained throughout life.

Although the paternal copy of *Dlk1* is typically the one expressed, mice harbouring the non-functional *Dlk1* copy of either maternal or paternal origin developed the neurogenic phenotypes observed in mice lacking both copies of the gene. In search of an explanation for this, the authors demonstrate that, as expected, differentiated neurons in non-neurogenic tissues expressed the paternal copy of *Dlk1*. However, NSCs and their niche astrocytes — cells that are similar to NSCs and are found in the NSC microenvironment — expressed both the maternal and paternal copies of *Dlk1* in P7 and adult animals (Fig. 1). An absence of imprinting in the neurogenic niche is a novel and unexpected concept, because it underscores an underappreciated functional link between epigenetic regulation, in the form of gene imprinting (and the removal thereof), and developmental events such as adult NSC homeostasis.

The precise function of *Dlk1* during neurogenesis remains unclear, but Ferrón and colleagues provide some early clues. The DLK1 protein comes in two forms: secreted and membrane-bound. Whereas niche astrocytes preferentially express secreted DLK1, NSCs predominantly express the membrane-bound isoform<sup>3</sup>. Significantly, loss of membrane-bound DLK1 from NSCs attenuates neurosphere formation in response to soluble DLK1, suggesting that secreted DLK1 stimulates the membrane-bound isoform in NSCs. Previous studies intimated<sup>10</sup> that DLK1 can bind to itself and activate a signal-transduction cascade that is independent of Notch signalling — the pathway in which DLK1 is thought to function. The nature of these alternative signalling cascades is tentative. But at least stem-cell biologists can now differentiate NSCs and niche astrocytes using two additional markers — membrane-bound and secreted DLK1.

This paper<sup>3</sup> raises many intriguing questions concerning the role of genomic imprinting and gene dosage during neurogenesis, as well as the potential mechanisms that regulate imprinting within a single organ. For example, do the same principles that underlie regulation of NSC development in the SVZ apply to the subgranular zone? And, if so, does *Dlk1* loss



**Figure 1 | *Dlk1* and development.** Ferrón *et al.*<sup>3</sup> find that *Dlk1* does not behave like a typical imprinted gene. SVZ, subventricular zone.



affect learning and memory? Studies involving tissue-specific and developmental-stage-specific loss of *Dlk1* function, as well as behavioural tests, should provide further evidence to support the current study.

More broadly, genomic imprinting serves as an important interface between environment and genes in mature tissues. Determining how *Dlk1* and other imprinted genes operate in concert with signalling cues, as well as discovering the factors that initiate and maintain differential methylation at the genomic region housing *Dlk1*, will lead to a better understanding of adult neurogenesis and brain development. Such knowledge will also inform emerging hypotheses about the role of trans-generational effects (or of their absence in particular processes), in which DNA and histone modifications in the ancestral genetic material affect the phenotypes of subsequent generations<sup>11</sup>. ■

**Edwin C. Oh and Nicholas Katsanis** are at the Center for Human Disease Modeling, Duke University, Durham, North Carolina 27710, USA. Nicholas Katsanis is also in the Departments of Cell Biology and Pediatrics. e-mail: katsanis@cellbio.duke.edu

1. Zhao, C. *et al.* *Cell* **132**, 645–660 (2008).
2. Alvarez-Buylla, A. & Lim, D. A. *Neuron* **41**, 683–686 (2004).
3. Ferrón, S. R. *et al.* *Nature* **475**, 381–385 (2011).
4. Reik, W. & Walter, J. *Nature Rev. Genet.* **2**, 21–32 (2001).
5. Wang, X. *et al.* *PLoS One* **3**, e3839 (2008).
6. Babak, T. *et al.* *Curr. Biol.* **18**, 1735–1741 (2008).
7. Meaney, M. J. & Ferguson-Smith, A. C. *Nature Neurosci.* **13**, 1313–1318 (2010).
8. da Rocha, S. T. *et al.* *Trends Genet.* **24**, 306–316 (2008).
9. Reynolds, B. A. & Weiss, S. *Science* **255**, 1707–1710 (1992).
10. Baladron, V. *et al.* *Front. Biosci.* **6**, A25–A32 (2001).
11. Nadeau, J. H. *Hum. Mol. Genet.* **18**, R202–R210 (2009).

underwater acoustic sounding shows that the sea bottom in the epicentral area moved seaward by as much as 24 metres and was uplifted by about 3 metres. Therefore, slip along the plate interface at depth must have exceeded the 27-metre peak slip inferred from the GeoNet data<sup>1</sup>; it may have been even more than 50 metres, as suggested from the joint modelling<sup>3</sup> of the GeoNet data and sea-bottom pressure records of the tsunami waves. For comparison, this is about twice the peak slip determined for the giant earthquakes in Sumatra in 2004 ( $M_w$  9.4) and in Chile in 2010 ( $M_w$  9.0) — and larger than that estimated for the biggest earthquake ever recorded<sup>4</sup>, the  $M_w$ -9.5 event of 1960, which ruptured more than 1,000 km of the plate boundary off the coast of southern Chile.

Over the 15 years preceding the 2011 event, the GeoNet data<sup>5</sup> had revealed the slow accumulation of strain across Honshu, with the Pacific plate squeezing and dragging down the eastern edge of Honshu. We know, however, that the coast of Honshu is being uplifted in the long term, so a significant fraction of that ‘interseismic’ strain — strain accumulating between earthquakes — must be compensated by sudden episodes of uplift. The current model holds that interseismic strain on the upper plate is purely elastic, and is ‘recovered’ during seismic rupture of the plate interface, so that in the long run the upper plate does not deform<sup>6</sup>. This assumption provides a rationale to relate slip on the plate interface to interseismic strain on the upper plate. Where the plate interface is creeping, strain on the upper plate is negligible; but where the plate interface is locked, the upper plate is compressed and dragged down, building up elastic strain until it is released when the locked patches slip.

Several earlier studies adopted this assumption<sup>5,7,8</sup> and found that the measured strain

## EARTHQUAKES

# The lessons of Tohoku–Oki

**An exceptional data set documents surface deformation before, during and after the earthquake that struck northeastern Japan in March 2011. But models for assessing seismic and tsunami hazard remain inadequate. SEE LETTER P.373**

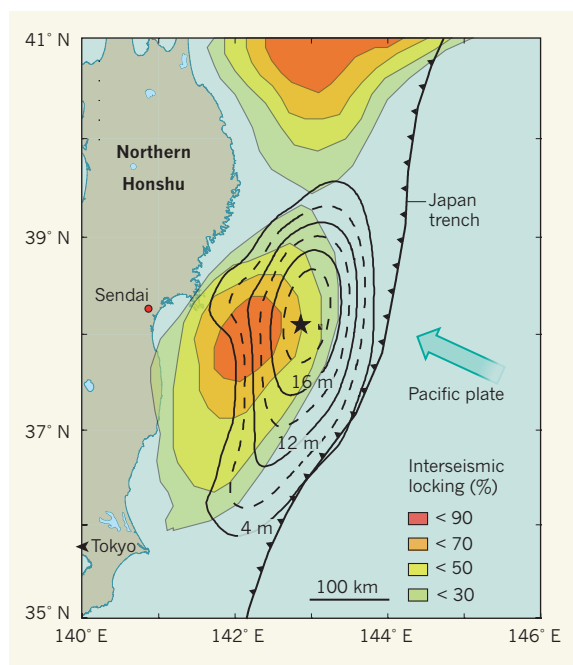
JEAN-PHILIPPE AVOUAC

Earthquake science has entered a new era with the development of space-based technologies to measure surface deformation at the boundaries of tectonic plates and large faults. Japan has been at the forefront in implementing these technologies, in particular with the deployment some 15 years ago of a network of continuously recording Global Positioning System (GPS) stations known as GeoNet. Papers analysing the data associated with the devastating Tohoku–Oki earthquake of 11 March 2011 are now appearing. One of these, by Ozawa *et al.*<sup>1</sup>, is published on page 373\* of this issue.

With a moment magnitude ( $M_w$ ) of 9.0, the Tohoku–Oki earthquake ranks among the largest ever recorded. The data collected at the GeoNet stations<sup>1</sup> indicate that it resulted from the sudden slip of a remarkably compact area (400 kilometres long by 200 kilometres wide) of the plate interface where the Pacific plate slides beneath the Okhotsk plate, on which northern Japan lies. The rupture area (Fig. 1) lies off the coastline of Honshu, Japan’s biggest island, and extends east nearly all the way to the Japan trench — hence the

particularly devastating tsunami produced by the earthquake.

Other new papers<sup>2,3</sup> provide further information. A combination<sup>2</sup> of GPS measurements and



**Figure 1 | Location of the Tohoku–Oki earthquake.**

The earthquake, with its epicentre marked by a star, ruptured the plate interface along which the Pacific plate slides beneath northern Honshu at a rate of 8 centimetres per year. Ozawa and colleagues’ analysis<sup>1</sup> shows that the rupture area and distribution of slip (represented by the black contour lines) roughly coincide with a patch of the plate interface that had remained locked over the preceding decades<sup>8</sup> (coloured area east of Sendai). The earthquake source was extremely compact and produced very large slip at relatively shallow depth (less than 20 km depth), hence the devastating tsunami. The other well-locked patch in the north coincides with rupture areas of large historical earthquakes (in particular, the  $M_w$ -8.5 Sanriku earthquake of 1896 and the  $M_w$ -8.2 Tokachi–Oki earthquake of 1968).

\*This article and the paper under discussion<sup>1</sup> were published online on 15 June 2011.

across Honshu required a large, locked patch off the coast of Sendai (Fig. 1). The rupture area of the Tohoku-Oki earthquake coincides quite well with that area<sup>3</sup>. A noteworthy discrepancy, however, is that the rupture reached closer to the Japan trench, where interseismic models suggested there was little locking. The particularly extensive, shallow slip seen in the Tohoku-Oki event could be due either to high preseismic stress left over from previous ruptures of the plate interface that failed to reach the trench, or, as seismological investigations suggest<sup>9</sup>, to specific properties of the plate interface. In any case, the observed slip requires the shallow plate interface to have remained locked, at least partially, in the period before the earthquake.

The published interseismic models<sup>5,7,8</sup> indicated little locking at shallow depth, essentially as a result of built-in methodological assumptions: the shallow portion of the plate interface is actually not well constrained if only onshore data are used<sup>1</sup>. These models may have been misleadingly interpreted as discounting the possibility of extensive shallow slip, which in fact occurred. Therefore, in the absence of direct constraints from sea-bottom geodesy, it may be preferable for models to assume maximum locking of the shallow zone of the plate interface. In fact, the interseismic data do not exclude a locked region off the coast of Sendai extending all the way to the shallow plate-boundary zone at the trench. Such an assumption raises questions for assessing the frequency of Tohoku-Oki-like earthquakes.

The estimated slip along the plate boundary off the coast of northern Honshu — due to earthquakes over the past few centuries — falls well short of balancing the slip deficit that should have accumulated over that period owing to interseismic locking. So it might seem that a large earthquake there was overdue. Indeed, according to the published interseismic models, interseismic strain builds up really fast on that boundary: it should take only a few centuries to accumulate enough strain to generate an  $M_w$  9.0 earthquake. Such large events should recur even more often if locking of the shallow portion of the plate interface is assumed in the modelling of interseismic strain.

By contrast, on the basis of historical and palaeo-tsunami records<sup>10</sup>, large earthquakes would be predicted to return only once every 1,000 years, or even less frequently. The way out of this conundrum is not clear. There is no evidence for particularly frequent episodes of large aseismic slip in that area, and postseismic afterslip, although significant<sup>1</sup>, is much too small to balance the slip budget. So either the slip deficit accumulating in the interseismic period is overestimated (which might happen if, for example, a fraction of interseismic strain is not recoverable), or it is incorrect to assume that geodetic rates measured over a decade or so are representative of strain build-up over periods of centuries to millennia.

Another paradoxical and possibly related observation is that the Tohoku-Oki earthquake induced more than 1 metre of systematic coastal subsidence, whereas uplift would have been expected to balance the subsidence rate of 5 millimetres per year during the interseismic period. The long-term coastal uplift requires deformation events that are large and frequent enough to compensate for that subsidence. This might call for a review of both the assumption that interseismic deformation of the upper plate is purely elastic, and the corollary that interseismic elastic strain is relaxed only by earthquakes that occur along the plate interface.

Finally, the geodetic data acquired both before<sup>5,7,8</sup> and after<sup>1</sup> the Tohoku-Oki earthquake suggest that the plate interface south of the rupture area is mostly creeping aseismically. There is thus no indication of a major zone of strain build-up on that portion of the plate boundary that might threaten Tokyo. But it is clear that although geodetic networks are invaluable instruments for observing strain accumulation and seismic release at plate

boundaries and major faults, we don't yet have an adequate theory to use these data for earthquake and tsunami hazard assessment. ■

**Jean-Philippe Avouac** is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

e-mail: avouac@gps.caltech.edu

1. Ozawa, S. *et al.* *Nature* **475**, 373–376 (2011).
2. Sato, M. *et al.* *Science* doi:10.1126/science.1207401 (2011).
3. Simons, M. *et al.* *Science* doi:10.1126/science.1206731 (2011).
4. Moreno, M. S. *et al.* *Geophys. Res. Lett.* **36**, L16310, doi:10.1029/2009gl039276 (2009).
5. Suwa, Y. *et al.* *J. Geophys. Res.* **111**, B04402, doi:10.1029/2004JB003203 (2006).
6. Savage, J. C. *Annu. Rev. Earth Planet. Sci.* **11**, 11–43 (1983).
7. Hashimoto, C. *et al.* *Nature Geosci.* **2**, 141–144 (2009).
8. Loveless, J. P. & Meade, B. J. *J. Geophys. Res. Solid Earth* **115**, B02410, doi:10.1029/2008jb006248 (2010).
9. Ide, S., Baltay, A. & Beroza, G. C. *Science* doi:10.1126/science.1207020 (2011).
10. Sawai, Y. *et al.* *Holocene* **18**, 517–528 (2008).

#### MICROBIOLOGY

## Molecular syringes scratch the surface

**New data suggest that the most recently discovered class of bacterial 'molecular syringes' inject proteins only across the outer membrane of target cells during interbacterial competition. SEE ARTICLE P.343**

PEGGY COTTER

Many bacteria use specialized secretion systems to inject proteins or DNA into cells of eukaryotic organisms (such as animals and plants) or into other bacteria. Little is known about the mechanism of secretion in the most recently discovered class of these molecular syringes, the type VI secretion system (T6SS)<sup>1</sup>. On page 343 of this issue, Mougous and colleagues (Russell *et al.*<sup>2</sup>) show that the T6SS of the bacterium *Pseudomonas aeruginosa* delivers two proteins into target bacteria. These proteins degrade peptidoglycan, a highly cross-linked lattice that lies just below the outer membrane of Gram-negative bacteria in a region called the periplasm, causing lysis of the target cell. These findings strongly suggest that the T6SS 'needles' puncture only one membrane (the bacterial outer membrane in this case), providing substantial insight into this system's mechanism of action.

T6SSs were discovered on the basis of their contribution to symbiosis and virulence in bacterial interactions with eukaryotes. Until recently, the only proteins that had been

shown to enter host cells through T6SSs were the VgrG proteins of the bacterium *Vibrio cholerae*, which seem to form the membrane-puncturing tip of the T6SS needle<sup>3</sup>.

Last year, Mougous and co-workers<sup>4</sup> identified three candidate T6SS-dependent 'effector' proteins in *P. aeruginosa*. One of these proteins, called Tse2, was toxic to both mammalian and bacterial cells if the cells were engineered to produce it intracellularly. The co-production of an immunity protein called Tsi2 prevented this toxicity. Surprisingly, however, *P. aeruginosa* itself could not intoxicate mammalian cells if co-cultured with them, but it could outcompete other Gram-negative bacteria in a manner dependent on both cell–cell contact and Tse2.

Subsequently, Pukatzki and colleagues<sup>5</sup> showed that *V. cholerae* could also outcompete other Gram-negative bacteria, and that it did so using the same T6SS that it uses to inject proteins into amoeba and mammalian macrophages. These results suggested that at least some T6SSs can deliver proteins into both mammalian cells and Gram-negative bacteria.

That T6SSs can inject proteins into both



eukaryotic cells and Gram-negative bacteria is remarkable, because the membrane structures of these cells are very different. Eukaryotic cells are bounded by a single phospholipid bilayer (the plasma membrane). By contrast, Gram-negative bacteria are bounded by two different phospholipid bilayers — the cytoplasmic membrane and the outer membrane, which contains lipopolysaccharides in its outer leaflet. The two membranes are separated by the periplasm, which contains the peptidoglycan lattice. So how could a single protein-delivery system function on such disparate substrates?

Mougous and colleagues<sup>2</sup> now characterize the other two candidate effectors that they discovered in their original screen<sup>4</sup> — Tse1 and Tse3. They demonstrate that Tse1 and Tse3 lyse target bacterial cells in a T6SS-dependent manner, that both are enzymes that degrade peptidoglycan, and that neither protein can access the periplasm when produced intracellularly or when added to the outside of intact bacterial cells. Moreover, they show that *P. aeruginosa* produces specific immunity proteins that protect it against Tse1- and Tse3-mediated interbacterial lysis, but only when these immune proteins are directed into the periplasm. The *P. aeruginosa* T6SS, therefore, seems to penetrate only the outer membrane, delivering effector proteins into the periplasm, and not directly into the cytoplasm, of target bacteria. T6SS-dependent delivery of proteins into eukaryotic cells and Gram-negative bacteria may therefore not be so different, in that it may require injection across only a single membrane in both cases.

These results provide functional support for

the hypothesis — based on structural similarities — that T6SSs operate similarly to bacterial viruses (bacteriophages)<sup>6,7</sup>. Bacteriophages of the Myoviridae family initiate infection by injecting their DNA into host bacteria using their long contractile tails<sup>8</sup>. These tails have a ‘tube within a tube’ structure in which the outer tube is ‘spring-loaded’ to contract when the proper host cell is recognized, forcing the rigid inner tube through the outer membrane and into the periplasm (Fig. 1a). The peptidoglycan-degrading activity of one of the membrane-puncturing tip proteins allows the tube to penetrate the peptidoglycan lattice. Contact between the inner tube and the cytoplasmic membrane then initiates DNA transfer into the cytoplasm, but the tube does not perforate the cytoplasmic membrane.

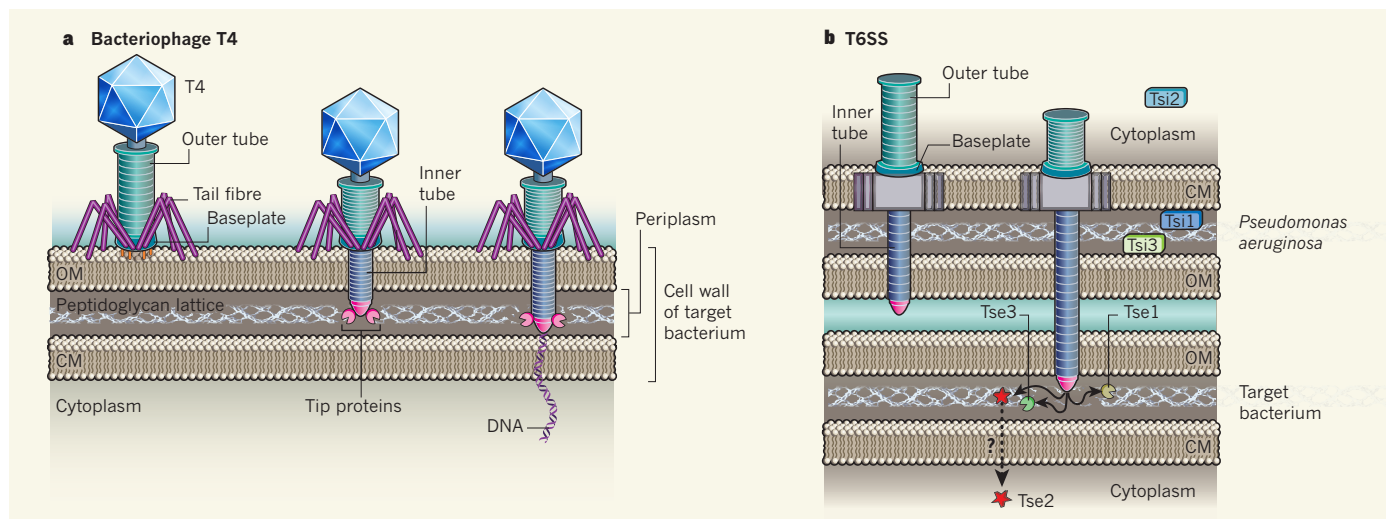
Mougous and colleagues’ results<sup>2</sup> indicate that, like the tail tube of bacteriophage T4, the needle of the *P. aeruginosa* T6SS is thrust only through the bacterial outer membrane (or the eukaryotic cell plasma membrane) when the outer tube contracts (Fig. 1b). The predicted membrane-puncturing tip proteins of T6SSs (VgrG proteins) lack obvious peptidoglycan-degrading activity, so the T6SS needle presumably cannot penetrate the peptidoglycan lattice. Although *P. aeruginosa* injects peptidoglycan-degrading effector proteins (Tse1 and Tse3), their function seems to be to kill the target cell rather than just to advance the needle. The authors also show that although a functional T6SS is required for Tse2-mediated interbacterial competition, Tse1 and Tse3 are not. This suggests that Tse2 is somehow translocated independently into the

cytoplasm, and that an intact peptidoglycan lattice does not prevent this movement.

Although Mougous and co-workers’ results shed light on how T6SSs function, many questions remain. For example, how is specificity for the target determined? Is it mediated by T6SS components or by separate proteins such as pili or other adhesins on the bacterial surface? In cases where a T6SS (such as the T6SS of *V. cholerae*) can inject different cell types, are different subsets of effector proteins involved? And if so, how is this controlled?

The biophysics of injection is also intriguing. In bacteriophage T4 infection, the phage’s baseplate binds to lipopolysaccharides of the host bacterium’s outer membrane, and this interaction is apparently stronger than the force required for the inner tail tube to puncture the outer membrane. For T6SSs, the baseplate seems to be located in the bacterial cytoplasm or possibly in the periplasm. What then holds the opposing membranes together at the injection site so that the T6SS needle punctures the membrane of the target cell rather than deforming either cell’s membrane or, perhaps, just pushing the cells apart?

Finally, the most difficult question may be why bacteria use contact-dependent delivery of proteins for interbacterial competition. Wouldn’t secreting antibacterial agents (such as antibiotics or bacteriocins) into the extracellular environment be a more efficient way to eliminate one’s neighbours? Does contact-dependence allow discrimination — to determine which bacteria to kill and which to spare? Is competition even the goal? Perhaps T6SS-dependent interbacterial interactions



**Figure 1 | Similar injection systems.** **a**, The bacteriophage T4 binds to the surface of a target bacterium with its tail fibres and baseplate, causing a conformational change in the outer-tube proteins and so outer-tube contraction. Consequently, the rigid inner tube is forced through the outer membrane (OM) of the bacterium. The peptidoglycan-degrading domains of the tip proteins then allow penetration through the periplasmic peptidoglycan lattice, and finally interaction with the cytoplasmic membrane (CM) initiates translocation of viral DNA into the cytoplasm. **b**, The bacterial type VI secretion system (T6SS) is proposed to function similarly, because several proteins in the

two systems are structurally similar (shown in the same colours)<sup>1,6,7</sup>. Mougous and colleagues<sup>2</sup> show that two T6SS effectors of *Pseudomonas aeruginosa* (Tse1 and Tse3) degrade peptidoglycan; their associated immunity proteins (Tsi1 and Tsi3) are protective only when located in the periplasm. These data support the model (shown) that T6SS delivers proteins across only one membrane, and that secreted proteins are delivered from the cytoplasm of the injecting cell into the periplasm of the target cell without entering the periplasm of the injecting cell. How Tse2, which seems to have a cytoplasmic target<sup>4</sup>, crosses the cytoplasmic membrane of a target bacterium is not known.

are involved in developmental processes such as formation of multicellular structures or of organized communities? Answering these questions will require a way to study T6SS-dependent behaviour of bacteria in their natural habitats. ■

**Peggy Cotter** is in the Department of Microbiology and Immunology, School of Medicine, University of North Carolina–Chapel Hill, Chapel Hill, North Carolina 27599, USA.

e-mail: [pcotter@med.unc.edu](mailto:pcotter@med.unc.edu)

1. Records, A. R. *Mol. Plant Microbe Interact.* **24**, 751–757 (2011).
2. Russell, A. B. *et al. Nature* **475**, 343–347 (2011).
3. Ma, A. T. *et al. Cell Host Microbe* **5**, 234–243 (2009).
4. Hood, R. D. *et al. Cell Host Microbe* **7**, 25–37 (2010).
5. MacIntyre, D. L. *et al. Proc. Natl Acad. Sci. USA* **107**, 19520–19524 (2010).
6. Kanamaru, S. *Proc. Natl Acad. Sci. USA* **106**, 4067–4068 (2009).
7. Bonemann, G. *et al. EMBO J.* **28**, 315–325 (2009).
8. Rossmann, M. G. *et al. Curr. Opin. Struct. Biol.* **14**, 171–180 (2004).

## ASTROPARTICLE PHYSICS

# A multi-messenger story

**The IceCube detector is a super-sensitive tool with which astronomers hope to find the elusive neutrinos from cosmic  $\gamma$ -ray bursts. A search that, surprisingly, has come up empty handed prompts a rethink of the underlying theory.**

DIETER H. HARTMANN

For some time now, basic ideas about the origin of cosmic  $\gamma$ -ray bursts (GRBs) have converged on the formation of massive, rapidly spinning black holes. This hypothesis is supported by observational evidence derived from the bursts' electromagnetic signatures<sup>1</sup>. But GRBs, which come in two classes (long and short), are also expected to generate other multi-messenger signatures — for example, gravity waves, cosmic rays and neutrinos. The last of these has been the focus of intensive searches; however, discoveries have yet to be reported. Writing in *Physical Review Letters*, Abbasi *et al.*<sup>2</sup> (the IceCube Collaboration) report that the search for GRB neutrinos with the IceCube detector, located at the South Pole, has come up empty handed. Carl Sagan's antimetabole "absence of evidence is not evidence of absence" immediately comes to mind: not detecting an expected signal can be as powerful a diagnostic as the arguably more desirable discovery, and may send us back to re-evaluate how we arrived at our expectations in the first place.

For many years, progress in neutrino astronomy was driven by the absence of a large fraction of the predicted solar neutrino flux. Interestingly, this puzzle was not solved by making modifications to the standard solar model, but instead by particle physics. The detection of a few neutrinos from the famous supernova SN 1987A in the nearby Large Magellanic Cloud galaxy established extragalactic neutrino astronomy. Even the diffuse neutrino background due to neutrinos from all supernovae throughout the Universe is now within reach of the Super-Kamiokande facility in Japan<sup>3</sup>.

Like supernovae, GRBs involve an incredible

release of energy, and they also partition their energy budget such that more than 99% of the energy is carried away by neutrinos. However, unlike their electromagnetic signal, the neutrino signal is hard to detect because of the large cosmological distances involved and the small interaction cross-section with matter.

The IceCube detector is the largest tool with which astronomers hope to catch the elusive neutrino signal from GRBs and other sources. A volume of about 1 cubic kilometre of ice is monitored by chains of light detectors buried deep in ancient ice layers. Abbasi and colleagues' search, which covered 117 GRBs recorded during about a year of operation, did not yield detections either coincident with GRBs or within the 24 hours following a GRB. This is surprising, because the detector's current configuration of 40 strings (about half its final design size) had finally reached the sensitivity at which the detection of neutrinos from GRBs was anticipated from current theoretical GRB models. The parallel to the puzzle of the missing solar neutrinos is obvious, but this time it seems more likely that the model builders will need to return to their drawing boards.

Should discovery really have happened at this point, or was it just an exciting possibility? To address this question, another astrophysical puzzle should be considered. Cosmic rays (CRs) — energetic particles (protons, helium nuclei and heavier species) filling space near the Sun and the Galaxy as a whole — are believed to be accelerated in the shock environments of supernovae. But their highest-energy component, reaching more than  $10^{20}$  electronvolts (eV), cannot be explained in this way. The mystery of such ultra-high-energy cosmic rays (UHECRs) may be linked to GRBs. One

possibility is that the high-energy tail of the particles' observed energy distribution function (above the so-called ankle at  $4 \times 10^{18}$  eV) is extragalactic in origin. GRBs have been proposed<sup>4,5</sup> as a powerful accelerator of CRs. Because CRs are deflected by magnetic fields, a direct connection between CRs and GRBs is almost impossible to establish. But there is an indirect link that could serve as 'proof'.

Waxman and Bahcall<sup>6</sup> pointed out that CRs escaping GRB acceleration sites would create neutrinos of high energy through interactions with the intense photon background in which they are immersed. Typical energies of neutrinos created during black-hole formation are of the order of 10 MeV, whereas those resulting from proton- $\gamma$ -ray interactions in GRBs fall in the TeV–PeV domain, which is, in principle, easier to detect. The proton- $\gamma$ -ray process leads to neutral and charged subatomic particles known as pions. Neutral pions decay to photons, whereas charged pions decay to a mix of neutrinos and charged particles known as muons, which, in turn, decay to yield more neutrinos.

The expected neutrino luminosity of a GRB depends on the relative portioning of the burst's energy, with the common assumption of protons and photons carrying equal amounts (50/50 energy-partition rule), and an uncertain fraction of 10–30% invoked for energy transfer into the charged-pion neutrino-production channel. The possibility of UHECR production by GRBs offers a solution to the energy puzzle well above the 'knee', where supernova-remnant shocks cannot reach, but this solution implies that copious amounts of neutrinos are co-produced. This was another reason for the high expectations of a successful IceCube search.

The predicted neutrino fluxes for each GRB searched for in the IceCube data were based on detailed model calculations by Guetta and co-workers<sup>7</sup>. Using a typical value of 20% for the energy transfer into the charged-pion neutrino-production channel and adopting the 50/50 energy-partition rule, one can quickly obtain a rough estimate of the number of neutrinos arriving at Earth per unit area. Typical time-integrated  $\gamma$ -ray fluxes of GRBs are of order of  $10^{-5}$  ergs cm<sup>-2</sup>, so that typical neutrino energies of about  $10^{15}$  eV correspond to about 10 neutrinos per square kilometre.

IceCube is a TeV-scale neutrino telescope, which predominantly sees neutrinos through the Cherenkov light from secondary muons that result from interactions between neutrinos and nucleons (protons or neutrons) in ice. The long mean free path of muons in ice (kilometres) allows a large volume to be monitored, and thus creates a sensitive detector for neutrinos in the TeV–PeV regime. However, the conversion probability of a neutrino to a muon within the range of the detector is much smaller than 100%, and observers therefore deal with small-number statistics in addition to systematic effects and background issues.



are involved in developmental processes such as formation of multicellular structures or of organized communities? Answering these questions will require a way to study T6SS-dependent behaviour of bacteria in their natural habitats. ■

**Peggy Cotter** is in the Department of Microbiology and Immunology, School of Medicine, University of North Carolina–Chapel Hill, Chapel Hill, North Carolina 27599, USA.

e-mail: [pcotter@med.unc.edu](mailto:pcotter@med.unc.edu)

1. Records, A. R. *Mol. Plant Microbe Interact.* **24**, 751–757 (2011).
2. Russell, A. B. *et al. Nature* **475**, 343–347 (2011).
3. Ma, A. T. *et al. Cell Host Microbe* **5**, 234–243 (2009).
4. Hood, R. D. *et al. Cell Host Microbe* **7**, 25–37 (2010).
5. MacIntyre, D. L. *et al. Proc. Natl Acad. Sci. USA* **107**, 19520–19524 (2010).
6. Kanamaru, S. *Proc. Natl Acad. Sci. USA* **106**, 4067–4068 (2009).
7. Bonemann, G. *et al. EMBO J.* **28**, 315–325 (2009).
8. Rossmann, M. G. *et al. Curr. Opin. Struct. Biol.* **14**, 171–180 (2004).

## ASTROPARTICLE PHYSICS

# A multi-messenger story

**The IceCube detector is a super-sensitive tool with which astronomers hope to find the elusive neutrinos from cosmic  $\gamma$ -ray bursts. A search that, surprisingly, has come up empty handed prompts a rethink of the underlying theory.**

DIETER H. HARTMANN

For some time now, basic ideas about the origin of cosmic  $\gamma$ -ray bursts (GRBs) have converged on the formation of massive, rapidly spinning black holes. This hypothesis is supported by observational evidence derived from the bursts' electromagnetic signatures<sup>1</sup>. But GRBs, which come in two classes (long and short), are also expected to generate other multi-messenger signatures — for example, gravity waves, cosmic rays and neutrinos. The last of these has been the focus of intensive searches; however, discoveries have yet to be reported. Writing in *Physical Review Letters*, Abbasi *et al.*<sup>2</sup> (the IceCube Collaboration) report that the search for GRB neutrinos with the IceCube detector, located at the South Pole, has come up empty handed. Carl Sagan's antimetabole "absence of evidence is not evidence of absence" immediately comes to mind: not detecting an expected signal can be as powerful a diagnostic as the arguably more desirable discovery, and may send us back to re-evaluate how we arrived at our expectations in the first place.

For many years, progress in neutrino astronomy was driven by the absence of a large fraction of the predicted solar neutrino flux. Interestingly, this puzzle was not solved by making modifications to the standard solar model, but instead by particle physics. The detection of a few neutrinos from the famous supernova SN 1987A in the nearby Large Magellanic Cloud galaxy established extragalactic neutrino astronomy. Even the diffuse neutrino background due to neutrinos from all supernovae throughout the Universe is now within reach of the Super-Kamiokande facility in Japan<sup>3</sup>.

Like supernovae, GRBs involve an incredible

release of energy, and they also partition their energy budget such that more than 99% of the energy is carried away by neutrinos. However, unlike their electromagnetic signal, the neutrino signal is hard to detect because of the large cosmological distances involved and the small interaction cross-section with matter.

The IceCube detector is the largest tool with which astronomers hope to catch the elusive neutrino signal from GRBs and other sources. A volume of about 1 cubic kilometre of ice is monitored by chains of light detectors buried deep in ancient ice layers. Abbasi and colleagues' search, which covered 117 GRBs recorded during about a year of operation, did not yield detections either coincident with GRBs or within the 24 hours following a GRB. This is surprising, because the detector's current configuration of 40 strings (about half its final design size) had finally reached the sensitivity at which the detection of neutrinos from GRBs was anticipated from current theoretical GRB models. The parallel to the puzzle of the missing solar neutrinos is obvious, but this time it seems more likely that the model builders will need to return to their drawing boards.

Should discovery really have happened at this point, or was it just an exciting possibility? To address this question, another astrophysical puzzle should be considered. Cosmic rays (CRs) — energetic particles (protons, helium nuclei and heavier species) filling space near the Sun and the Galaxy as a whole — are believed to be accelerated in the shock environments of supernovae. But their highest-energy component, reaching more than  $10^{20}$  electronvolts (eV), cannot be explained in this way. The mystery of such ultra-high-energy cosmic rays (UHECRs) may be linked to GRBs. One

possibility is that the high-energy tail of the particles' observed energy distribution function (above the so-called ankle at  $4 \times 10^{18}$  eV) is extragalactic in origin. GRBs have been proposed<sup>4,5</sup> as a powerful accelerator of CRs. Because CRs are deflected by magnetic fields, a direct connection between CRs and GRBs is almost impossible to establish. But there is an indirect link that could serve as 'proof'.

Waxman and Bahcall<sup>6</sup> pointed out that CRs escaping GRB acceleration sites would create neutrinos of high energy through interactions with the intense photon background in which they are immersed. Typical energies of neutrinos created during black-hole formation are of the order of 10 MeV, whereas those resulting from proton- $\gamma$ -ray interactions in GRBs fall in the TeV–PeV domain, which is, in principle, easier to detect. The proton- $\gamma$ -ray process leads to neutral and charged subatomic particles known as pions. Neutral pions decay to photons, whereas charged pions decay to a mix of neutrinos and charged particles known as muons, which, in turn, decay to yield more neutrinos.

The expected neutrino luminosity of a GRB depends on the relative portioning of the burst's energy, with the common assumption of protons and photons carrying equal amounts (50/50 energy-partition rule), and an uncertain fraction of 10–30% invoked for energy transfer into the charged-pion neutrino-production channel. The possibility of UHECR production by GRBs offers a solution to the energy puzzle well above the 'knee', where supernova-remnant shocks cannot reach, but this solution implies that copious amounts of neutrinos are co-produced. This was another reason for the high expectations of a successful IceCube search.

The predicted neutrino fluxes for each GRB searched for in the IceCube data were based on detailed model calculations by Guetta and co-workers<sup>7</sup>. Using a typical value of 20% for the energy transfer into the charged-pion neutrino-production channel and adopting the 50/50 energy-partition rule, one can quickly obtain a rough estimate of the number of neutrinos arriving at Earth per unit area. Typical time-integrated  $\gamma$ -ray fluxes of GRBs are of order of  $10^{-5}$  ergs cm<sup>-2</sup>, so that typical neutrino energies of about  $10^{15}$  eV correspond to about 10 neutrinos per square kilometre.

IceCube is a TeV-scale neutrino telescope, which predominantly sees neutrinos through the Cherenkov light from secondary muons that result from interactions between neutrinos and nucleons (protons or neutrons) in ice. The long mean free path of muons in ice (kilometres) allows a large volume to be monitored, and thus creates a sensitive detector for neutrinos in the TeV–PeV regime. However, the conversion probability of a neutrino to a muon within the range of the detector is much smaller than 100%, and observers therefore deal with small-number statistics in addition to systematic effects and background issues.

The temporal coincidence of neutrino counts with photons from GRBs is of great help for background reduction.

Abbasi *et al.*<sup>2</sup> report on model-dependent searches, in which specific GRB models were applied to identify prompt neutrino emission, and model-independent searches, in which wider time windows (up to a day) were used and GRB specifics were not assumed. Neither approach found the elusive neutrino signal. In another study<sup>8</sup>, the IceCube Collaboration analysed a sample of 36,900 astrophysical objects in terms of an all-sky map, and no statistically significant neutrino signal emerged. Neutrino astronomy is a challenging field, but a breakthrough may be just around the corner.

The production of TeV neutrinos from GRBs is connected to the idea that UHECRs are produced and released in GRBs as well (for a recent review see ref. 9). The story of the missing TeV neutrinos from GRBs is thus an excellent example of the emerging field of multi-messenger astrophysics. Cosmic ray,  $\gamma$ -ray and neutrino astronomy are closely connected in this story. Absence of evidence may not be evidence for absence (of TeV neutrinos from GRBs), but the fact remains that our expectations were not fulfilled, and that we are now forced to reconsider assumptions about the physics of these sources.

Recent evidence<sup>10</sup> for an emerging class of very energetic bursts (a tenfold larger energy output than the canonical  $10^{51}$  ergs) — exemplified by GRB 090926A, which was detected by the Fermi space observatory — suggests that the GRB models based on the formation of rapidly spinning black holes may have to be augmented. And perhaps our understanding of this class will be aided by future neutrino detections. Theorists may tune existing models or invent new ones, but if the fully developed IceCube detector — with twice its current set of strings — still produces only upper limits for GRB–neutrino associations, it will be time to reconsider the hypothetical GRB origin of UHECRs. Discovery of TeV neutrinos from GRBs would have been spectacular, but even constraints drive the development of this multi-messenger puzzle. ■

**Dieter H. Hartmann** is in the Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634, USA.  
e-mail: hdieter@clemson.edu

## MOLECULAR PROGRAMMING

# DNA and the brain

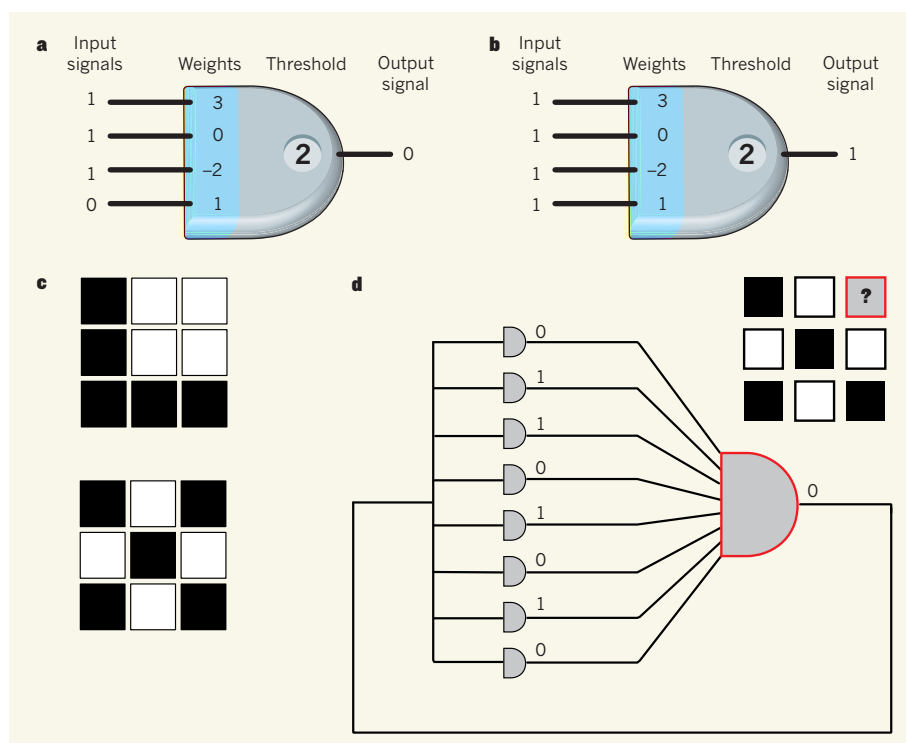
**The idea that artificial neural networks could be based on molecular components is not new, but making such a system has been difficult. A network of four artificial neurons made from DNA has now been created. SEE LETTER P.368**

ANNE CONDON

The design of intelligent systems is a long-standing goal for scientists, not least those in the Acme Labs of the animated TV series *Pinky and the Brain*. The Acme researchers used their technology to enhance the intelligence of the eponymous mice — Brain became a fiendish genius bent on world domination, although Pinky's transformation into a dimwit was arguably less impressive. Such experiments are clearly fantasy, but a related and compelling bioengineering challenge in the real world is to demonstrate

how tiny biological molecules could support limited forms of intelligent behaviour, as must have happened before brains evolved. On page 368 of this issue, Qian *et al.*<sup>1</sup> report a leap forward in this area: a network of interacting DNA strands that can act as artificial neurons, and that supports simple memory functions.

Brains are large networks of neurons. Within these networks, individual cells produce electrochemical signals whose strength depends in a complex way on the strengths of input signals received from other neurons in the network, or from sensory inputs. Artificial neurons are theoretical, highly simplified



**Figure 1 | Pattern recognition by artificial neural networks.** **a**, The diagram represents an artificial neuron that has four inputs and one output. If the inputs from top to bottom are 1, 1, 1 and 0, then the weighted sum of inputs is  $3 + 0 - 2 + 0 = 1$ . This is less than the threshold of 2, and so the output is 0. **b**, If the inputs to the same neuron are all 1, then the weighted sum of inputs is 2, and the output is 1. **c**, Networks of artificial neurons can be used for pattern recognition. Here, the letters L and X are depicted as patterns of nine black and white squares in a grid. **d**, A network of nine artificial neurons, where each neuron corresponds to a square in the grid, can identify whether an incomplete pattern, such as that shown, is L or X. Each neuron receives signals from all the other neurons, but, for simplicity, only the signals to and from the neuron associated with the top-right square — the large red neuron in the diagram — are shown. Neurons associated with white squares provide input values of 1, whereas those associated with black squares provide a value of 0. On the basis of its predetermined weightings and threshold value (not shown), the red neuron determines that the signal from the top-right square is 0 — that is, the square is black. Qian *et al.*<sup>1</sup> have made a DNA-based network of four artificial neurons that distinguishes between four four-bit patterns, and that reconstructs the patterns on the basis of incomplete descriptions.

1. Vedrenne, G. & Atteia, J.-L. *Gamma Ray Bursts* (Springer, 2009).
2. Abbasi, R. *et al. Phys. Rev. Lett.* **106**, 141101 (2011).
3. Horiuchi, S., Beacom, J. F. & Dwek, E. *Phys. Rev. D* **79**, 083013 (2009).
4. Vietri, M. *Astrophys. J.* **453**, 883–889 (1995).
5. Waxman, E. *Phys. Rev. Lett.* **75**, 386–389 (1995).
6. Waxman, E. & Bahcall, J. *Phys. Rev. Lett.* **78**, 2292–2295 (1997).
7. Guetta, D. *et al. Astropart. Phys.* **20**, 429–455 (2004).
8. Abbasi, R. *et al. Astrophys. J.* **732**, 18 (2011).
9. Mészáros, P. & Razzaque, S. in *Energy Budget in the High Energy Universe* (eds Sato, K. & Hisano, J.) 84–93 (World Scientific, 2007).
10. Cenko, S. B. *et al. Astrophys. J.* **732**, 29 (2011).



models of neurons<sup>2</sup> that produce a signal if the weighted sum of their inputs exceeds a threshold value (Fig. 1a,b). Because of their simplicity, networks of artificial neurons are but a shadow of the means used for information processing in the brain. Nevertheless, artificial neural networks implemented computationally are adept at pattern-association tasks that our brains do well, such as identifying letters of the alphabet in poor handwriting.

To understand how artificial neural networks perform such tasks, consider a pair of simple patterns:  $3 \times 3$  grids of black or white squares that represent two letters in the alphabet (Fig. 1c). Given an incomplete description of a pattern, artificial neural networks use an automated method to find the letter that best matches it. The nine inputs to such a network describe the incomplete pattern, with black squares represented by '0' and white squares by '1'. Squares whose colour is unknown are represented by '?' (Fig. 1d). The nine outputs of the network should describe the pattern that best matches the incomplete input, using '0's and '1's as above, '?' for squares whose colour couldn't be resolved and, in some cases, 'x' if the input is invalid.

The network's agents are nine artificial neurons, each of which corresponds to a square on the grid. Each neuron determines one of the nine outputs, using signals from all the other neurons as clues. Roughly, the weighted-threshold feature of an artificial neuron provides a sort of voting mechanism for its incoming 1-valued signals. For example, the middle-left and bottom-middle squares in Figure 1d are white (1-valued), which implies that the top-right square should be black (0-valued). Accordingly, in the neuron that corresponds to the top-right square, the inputs from the two white squares should be weighted to help bring the overall sum of inputs below the threshold value of the neuron, thus ensuring that the output is '0'. In other words, the inputs from the two white squares should be negative numbers (or negative votes).

By contrast, if a different incomplete pattern had a white square in the centre of the grid, the centre square's input signal to the 'top-right' neuron should be positively weighted, helping to ensure an output of '1' to indicate that the top-right square is white. The weights used by each neuron are determined in advance from a collection of patterns — that is, before any incomplete pattern is provided. In effect, the weights are a neuron's means of 'remembering' the collection of patterns, enabling the neuron to match incomplete patterns.

To date, efforts to synthesize molecular systems that behave as artificial neurons have been on too small a scale to mimic the action of a single neuron. But Qian *et al.*<sup>1</sup> have now built a network of four artificial neurons that distinguishes between four four-bit patterns, and that can identify which of these patterns matches an incomplete description. Their

network is built entirely from DNA.

The authors constructed their artificial neurons from modules that add, multiply and compute thresholds. These arithmetic modules were in turn built from more primitive subcomponents called see-saw gates — versatile units that two of the authors had previously used<sup>3</sup> in a quite different demonstration of digital logic circuits. The gates use different concentrations of two designated DNA strands to represent the three possible values of a signal: a high concentration of the first strand signals '0'; a high concentration of the second strand signals '1'; low concentrations of both strands signal '?'; and high concentrations of both strands signal 'x', indicating that the input does not match any pattern. Combinations of input DNA strands that are present in sufficiently high concentrations are converted by the see-saw gates into high concentrations of different output DNA strands, which in turn can be fed as input into other gates.

At the molecular level, see-saw gates use DNA-strand displacement as the basis of their function. Strand displacement happens when single-stranded DNA used as input forms duplexes with complementary strands in stable, multi-stranded complexes. The formation of new duplexes displaces extant strands of the original complex, which act as output.

Although Qian and colleagues' demonstration<sup>1</sup> of an artificial neural network is technically impressive, its small scale and computing power are, alas, more reminiscent of Pinky than of the Brain. Another limitation is that the

neuronal weights of the systems — in effect, the memory of the network — were predetermined using computer simulations, and are fixed. By contrast, our brains improve their performance in memory-association tasks, such as handwriting recognition, by fine-tuning the strengths of neuronal connections.

Nevertheless, the authors' DNA-based network is exciting because it shows how a biochemical system can remember information, and can use its memory to adapt to a changing environment by adjusting chemical concentrations. Because the network is built from a nucleic acid, it also provides a possible model for precursors of brains that existed in the RNA world — a postulated era of Earth in which all life was based on RNA molecules, rather than DNA. Moreover, the work opens the door to the development of biochemical neural networks that could fine-tune their neuronal weights over time, given appropriate feedback. In other words, it might pave the way for biochemical systems that can learn. ■

Anne Condon is in the Department of Computer Science, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.  
e-mail: [condon@cs.ubc.ca](mailto:condon@cs.ubc.ca)

1. Qian, L., Winfree, E. & Bruck, J. *Nature* **475**, 368–372 (2011).
2. McCulloch, W. S. & Pitts, W. *Bull. Math. Biophys.* **5**, 115–133 (1943).
3. Qian, L. & Winfree, E. *Science* **332**, 1196–1201 (2011).

#### STRUCTURAL BIOLOGY

## Peering into the spark of life

**Sodium channels in cell membranes have a crucial role in triggering bioelectrical events that lead to processes such as muscle contraction or hormone release. A crystal structure reveals how one such channel might work. [SEE ARTICLE P.353](#)**

RICHARD HORN

In 1786, Luigi Galvani famously observed the twitch of dissected frog legs in response to an electrical spark generated during a thunderstorm. What he didn't realize was that when a frog simply feels like jumping, the idea itself begins with a 'spark' — a bioelectrical event. We now know that these events are action potentials caused by a brief influx of positively charged sodium ions into excitable cells, such as neurons and muscle cells<sup>1</sup>. We also know that the influx of ions is gated by membrane proteins called sodium channels. But, despite more than 50 years of speculation

and intense experimentation, the structure of these proteins was unknown. Now, on page 353 of this issue, Payandeh *et al.*<sup>2</sup> report the crystal structure of the sodium channel NavAb from the bacterium *Arcobacter butzleri*, allowing us to peer inside the protein and see how these ion channels might work.

Sodium channels are members of a large class of voltage-gated ion channels (VGICs) that also includes potassium and calcium channels. They have a special status among VGICs, because almost all action potentials in vertebrates are initiated and caused by the transitory opening of sodium channels in response to a change of potential across the

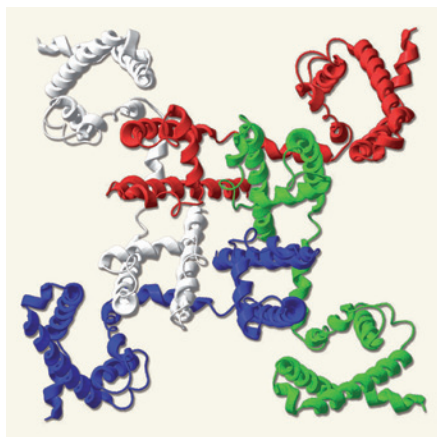
plasma membrane of an excitable cell. Sodium channels distinguish themselves from other VGICs not only in their selectivity for sodium ions, but also in the blazing speed with which they open in response to a stimulus. Another characteristic is that open sodium channels inactivate rapidly if the activating stimulus is maintained. The activation and inactivation gates of sodium channels, either of which may interrupt ion flux, are likely to be located in different parts of the protein<sup>3,4</sup>.

Like many other VGICs, sodium channels are exquisitely sensitive to small changes of membrane potential — a depolarization as small as 10 millivolts can increase the probability of their being open by a hundredfold<sup>5</sup>. Because the concentration of sodium ions outside cells is typically ten times that inside, channel opening causes a passive influx of positively charged sodium ions; the resulting depolarization induces more sodium channels to open. This positive-feedback cycle underpins the avalanche-like nature of action potentials, and explains why they can propagate for long distances along the surface of an excitable cell. Action potentials lead inexorably to one of two outcomes, depending on the cell type: the release of a molecule such as a neurotransmitter or hormone, or the contraction of a muscle cell.

VGICs are made up either of four identical subunits, as NavAb (Fig. 1)<sup>2</sup> and most potassium channels are, or from a long protein with four structurally similar (but not identical) subunit-like domains, as observed for sodium and calcium channels in animals. All VGICs contain a central pathway for ions and water that is surrounded by a tube known as the pore domain. This domain insulates the ions from the surrounding lipid membrane, thereby reducing the energy barrier to ions traversing the membrane. At the periphery of the pore domain are four voltage-sensing domains, one from each subunit (or subunit-like domain). Each voltage-sensing domain is composed of four transmembrane segments (S1–S4) connected together by three loops. The pore domain comprises the S5 and S6 segments of each subunit (or subunit-like domain), with each S5–S6 pair connected by an intervening loop (see Fig. 1a of the paper<sup>2</sup>).

The voltage-sensing domains detect changes in membrane potential largely through several positively charged arginine amino-acid residues in the S4 segment. These arginine side chains pull the associated S4 segment back and forth through the electric field across the membrane, in a process akin to a miniaturized form of electrophoresis. The S4 movement is coupled to the activation gate at the intracellular end of the permeation pathway.

In previously reported crystal structures of voltage-gated potassium channels<sup>6,7</sup>, the activation gate was open. However, in Payandeh and colleagues' structure<sup>2</sup> the activation gate of NavAb is shut tight. Interestingly, the authors observed that the S4 segments are in an outward,



**Figure 1 | Bird's-eye view of a voltage-gated ion channel.** Payandeh *et al.*<sup>2</sup> report the crystal structure of NavAb, a voltage-gated sodium channel. Here, NavAb is seen from 'above', with the axis of the channel perpendicular to the page. The tetrameric architecture is representative of voltage-gated ion channels. Each subunit is shown in a different colour; the four peripheral 'petals' are the voltage-sensing domains, and the ion-permeation pathway is at the centre of the image.

activated conformation that should favour an open channel. Channel opening in VGICs is typically preceded by outward movement of the S4 segments, and so the authors suggest that their structure reveals NavAb in a 'pre-open' state along the activation pathway of the protein. Another possibility is that the structure represents an inactivated state if the activation gate also serves as the inactivation gate, as seen in hyperpolarization-activated ion channels<sup>8</sup>.

The structure of NavAb has several fascinating features, three of which I consider here. The first concerns the S4 segment. As in other VGICs, each of its arginine residues is separated from the next by two hydrophobic residues. More notably, in NavAb, part of the S4 segment is wound into a particularly tight helix known as a 3<sub>10</sub>-helix — something that has also been seen in two potassium channels<sup>7,9</sup>. Consequently, the four voltage-sensing arginine residues near the extracellular ends of the S4 segments of NavAb form a linear array of charged side chains along the S4 axis. This suggests that the S4 segment moves like a piston along its axis, rather than twisting like a helical screw, as was speculated in earlier models of voltage-dependent S4 movement<sup>10,11</sup>. However, the possibility that S4 undergoes dynamic changes to its secondary structure as it moves cannot be ruled out.

The second notable feature is NavAb's selectivity filter — the structure that allows the channel to select sodium ions, rather than other ions, for passage through the pore. The filter contains a ring of four glutamic acid residues, one from each subunit, which makes it strongly negatively charged. Such a high field-strength anionic site was predicted<sup>12</sup> long ago

for molecules that selectively bind sodium over potassium ions. Nevertheless, Payandeh and colleagues' structure doesn't close the book on sodium selectivity, because sodium channels in animals have two acidic, one basic and one neutral residue, instead of the four acidic residues of NavAb.

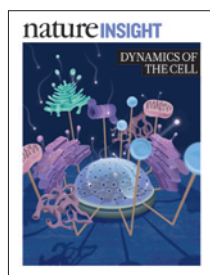
Perhaps the most intriguing features of the crystal structure<sup>2</sup>, however, are the fenestrations. The NavAb channel has four portals around it, midway through the transmembrane region (see Fig. 4 of the paper<sup>2</sup>). Such fenestrations are not found in potassium channels. Payandeh *et al.* observed that the acyl chains of lipid molecules extend through these windows into the middle of NavAb's ion-permeation pathway. As in potassium channels<sup>7,13</sup>, a phenylalanine residue near the middle of the inner helix of each subunit lining the pore may play a crucial part in ion permeation and gating. In NavAb, these residues sit in the fenestrations, where they can monitor and potentially participate in the movement of molecules between the surrounding membrane bilayer and the central cavity of the pore.

Sodium channels are known to be unusually sensitive to small pore-blocking molecules such as local anaesthetics and related compounds<sup>4,14</sup>, and these moderately hydrophobic molecules can somehow enter and exit the closed channels — but how they do this is a mystery. The fenestrations may be the explanation. If so, then the portals are ripe for further investigation, to find out how these clinically important drugs interact with the channel. The authors also suggest the tantalizing possibility that the fenestrations open and close depending on the gating state of the channel. Could the fenestrations open when the activation gate closes? Stay tuned! ■

**Richard Horn** is in the Department of Molecular Physiology and Biophysics, Jefferson Medical College, Philadelphia, Pennsylvania 19107, USA.  
e-mail: richard.horn@jefferson.edu

- Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer, 2001).
- Payandeh, J., Scheuer, T., Zheng, N. & Catterall, W. A. *Nature* **475**, 353–358 (2011).
- Bezanilla, F. & Stefani, E. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 819–846 (1994).
- Catterall, W. A. *Novartis Found. Symp.* **241**, 206–218 (2002).
- Hirschberg, B., Rovner, A., Lieberman, M. & Patlak, J. *J. Gen. Physiol.* **106**, 1053–1068 (1995).
- Long, S. B., Campbell, E. B. & MacKinnon, R. *Science* **309**, 903–908 (2005).
- Long, S. B. *et al. Nature* **450**, 376–382 (2007).
- Shin, K. S., Maertens, C., Proenza, C., Rothberg, B. S. & Yellen, G. *Neuron* **41**, 737–744 (2004).
- Clayton, G. M. *et al. Proc. Natl Acad. Sci. USA* **105**, 1511–1515 (2008).
- Catterall, W. A. *Annu. Rev. Biochem.* **55**, 953–985 (1986).
- Guy, H. R. & Seetharamulu, P. *Proc. Natl Acad. Sci. USA* **83**, 508–512 (1986).
- Eisenman, G. *Biophys. J.* **2**, 259–323 (1962).
- Cuello, L. G. *et al. Nature* **466**, 272–275 (2010).
- Fozzard, H. A., Lee, P. J. & Lipkind, G. M. *Curr. Pharm. Des.* **11**, 2671–2686 (2005).





Cover illustration by  
Nik Spencer

**Editor, *Nature***  
Philip Campbell

**Publishing**  
Nick Campbell

**Insights Editor**  
Ursula Weiss

**Production Editor**  
Nicola Bailey

**Senior Art Editor**  
Kelly Buckheit Krause

**Art Editor**  
Nik Spencer

**Sponsorship**  
Gerard Preston  
Yvette Smith

**Production**  
Emilia Orviss

**Marketing**  
Elena Woodstock  
Hannah Phipps

**Editorial Assistant**  
Hazel Mayhew

Advances in imaging in the past decade have revolutionized the field of cell biology. Developments such as super-resolution fluorescence microscopy and the ability to detect single molecules mean that molecular and organelle dynamics can now be visualized at very high temporal and spatial resolution within living cells, allowing the processes to be studied with unprecedented detail and precision.

It is becoming clear that the central dogma of molecular biology — DNA makes RNA makes protein — is overly simplistic. Gene-Wei Li and Sunney Xie look at how single-molecule microscopy has been used to monitor gene expression and regulation in real time, revealing the complex interactions within these processes.

In the spirit of a shift in perspectives on cellular dynamicity, Martin Schwartz and his colleagues explain why mechanotransduction — whereby the cell communicates with the external environment through mechanical signals — should not be considered a switch-like process. Instead, they argue, the subcellular structures that mediate mechanotransduction continually change their structure and composition in response to the varying forces they experience.

One of the most dynamic cellular processes is the folding of proteins into distinct three-dimensional structures. This process is regulated by a formidable network of proteins, and breaks down in various diseases and in ageing. Ulrich Hartl and his co-workers discuss the role of chaperones in protein folding and proteome maintenance, focusing on ways in which their substrate proteins navigate the complex folding-energy landscape of the cellular environment.

Cellular processes are kept running smoothly by orchestrated movements of macromolecules. For instance, complexes in the nuclear membrane mediate the exchange of molecules between the nucleus and the cytoplasm. Robert Singer and his colleagues present models for how RNA is exported from the nucleus, using evidence obtained through single-molecule imaging.

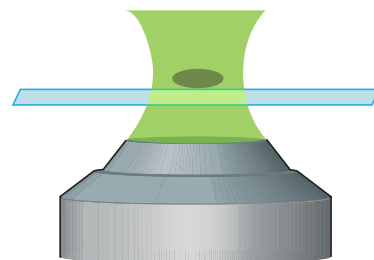
This Insight includes reviews on some of the most exciting advances in cell biology. As always, *Nature* carries sole responsibility for all editorial content and peer review.

**Deepa Nath**, Senior Editor  
**Sadaf Shadan**, Senior News & Views Editor

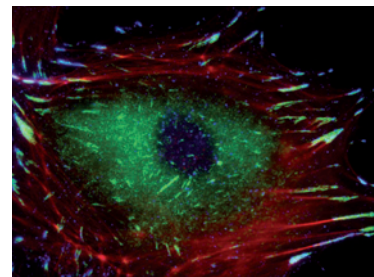
### CONTENTS

#### REVIEW

- 308 Central dogma at the single-molecule level in living cells**  
*G-W Li & X S Xie*

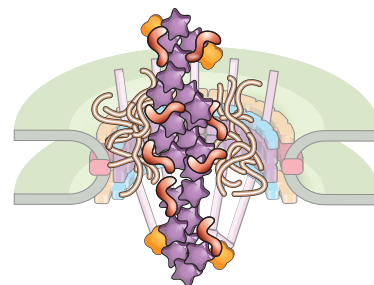


- 316 Dynamic molecular processes mediate cellular mechanotransduction**  
*B D Hoffman, C Grashoff & M A Schwartz*



- 324 Molecular chaperones in protein folding and proteostasis**  
*F U Hartl, A Bracher & M Hayer-Hartl*

- 333 Nuclear export dynamics of RNA-protein complexes**  
*D Grünwald, R H Singer & M Rout*



# Central dogma at the single-molecule level in living cells

Gene-Wei Li<sup>1</sup>† & X. Sunney Xie<sup>1</sup>

**Gene expression originates from individual DNA molecules within living cells. Like many single-molecule processes, gene expression and regulation are stochastic, that is, sporadic in time. This leads to heterogeneity in the messenger-RNA and protein copy numbers in a population of cells with identical genomes. With advanced single-cell fluorescence microscopy, it is now possible to quantify transcriptomes and proteomes with single-molecule sensitivity. Dynamic processes such as transcription-factor binding, transcription and translation can be monitored in real time, providing quantitative descriptions of the central dogma of molecular biology and the demonstration that a stochastic single-molecule event can determine the phenotype of a cell.**

This year marks the thirty-fifth anniversary of single-molecule optical detection and imaging. In 1976, Thomas Hirschfeld successfully detected single molecules at room temperature using an optical microscope to reduce probe volume and hence the background signal<sup>1</sup>. Figure 1a shows his one-dimensional (1D) fluorescence image of individual immobilized protein molecules, each labelled with tens of fluorophores. The use of tightly focused laser beams eventually allowed single-fluorophore detection in solution phase at room temperature, more than a decade later<sup>2</sup>. The imaging of single fluorophores in ambient environments was first reported with a scanning probe method<sup>3</sup>, and was followed by much easier and improved methods<sup>4–8</sup> akin to Hirschfeld's, which remain the methods of choice for imaging single molecules. In the past decade, improvements in photodetectors and optical components have allowed extensive single-molecule fluorescence studies on a variety of biological problems, first *in vitro* and more recently in living cells.

In a single-molecule experiment, one often observes stochastic behaviour, which would otherwise be obscured in an ensemble measurement. Figure 1b shows an early real-time observation of enzymatic turnovers of a single enzyme molecule, cholesterol oxidase<sup>9</sup>. The enzyme contains a flavin moiety that is naturally fluorescent in its oxidized form, but not in its reduced form. Each on/off cycle of fluorescence emission corresponds to an enzymatic turnover. This time trace resembles the electrical signal of a single ion channel recorded using a patch clamp — the first single-molecule technique in biology<sup>10</sup>. However, in this case, stochastic chemical reaction events of a single enzyme molecule are seen. Here, stochastic means that each fluorescence on/off time is probabilistic. Unlike the deterministic chemical kinetics of ensembles, each time trace is different, although their statistical properties are reproducible. On a single-molecule basis, when a chemical reaction occurs, a chemical bond is formed in less than 1 ps and the process cannot be resolved in a single-molecule experiment. However, the waiting time for the event is much longer and is probabilistic. When the kinetic scheme of a reaction includes a rate-limiting step, the distribution of the waiting times follows a single exponential distribution, and the number of events in a fixed time interval follows a Poisson distribution.

By contrast, if the overall reaction does not have one rate-limiting step but instead consists of identical sequential steps, the total waiting

time is less stochastic. An example of this is DNA replication by a single DNA polymerase, which is the basis of single-molecule sequencing<sup>11</sup>, a key application of single-molecule enzymology in biotechnology. A stochastic time trace of individual nucleotides incorporated into a single-stranded DNA template by a single DNA polymerase molecule is shown in Fig. 1c. Although the waiting time for each base incorporation step is exponentially distributed, the total waiting time for replicating the long DNA is narrowly distributed<sup>12</sup> — a consequence of the central limit theorem. Bacterial cell-cycle time, when limited by chromosome replication, is not stochastic for this reason<sup>13</sup>. The experiments in Fig. 1b, c were conducted under non-equilibrium steady-state conditions, in which the substrate concentration (thermodynamic driving force) does not change while substrate molecules are continuously converted to product molecules. This is similar to many non-equilibrium processes in a living cell, such as gene expression.

The central dogma of molecular biology states that genetic information encoded in DNA is transcribed to mRNA by RNA polymerases, and mRNA is translated to protein by ribosomes. In a living cell, DNA exists as individual molecules from which the regulation of gene expression originates. But our knowledge of gene expression has come mainly from genetic and biochemical studies conducted with large populations of cells and purified biomolecules, which often obscure the single-molecule nature of gene expression. In recent years, single-molecule experiments *in vitro* have provided mechanistic insight into the functions of macromolecules involved in gene expression, including transcriptional and translational machineries<sup>14–18</sup>. Compelling areas of further investigation involve the observation and quantitative description of gene expression and regulation in a living cell.

Not only is there only one copy (or a few copies) of a particular gene, but the copy number of a particular mRNA is also small owing to short intracellular mRNA lifetimes<sup>19</sup>, at least in a bacterial cell. The copy number of particular proteins ranges from zero to 10,000 (refs 20, 21); many important proteins, such as transcription factors, which regulate gene expression, have small copy numbers<sup>22</sup>. Consequently, single-molecule sensitivity for mRNA and protein is needed to quantify gene expression in individual cells.

Because of the stochasticity associated with the single or low copy-number macromolecules, the gene expression of individual cells cannot be synchronized. It is therefore necessary to make real-time

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. †Present address: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, USA.



observations of gene expression and regulation in single living cells. In particular, stochastic binding or unbinding of transcription factors to a particular gene, when rate limiting, must result in stochastic mRNA production, just like the single enzyme traces in Fig. 1b, c. Stochastic degradation of individual mRNA molecules further contributes to fluctuations in protein production. These temporal fluctuations of the mRNA and protein numbers (see sketches in Fig. 2) result in cell-to-cell variation of the copy numbers, or gene expression 'noise'. Under the steady-state condition, the connection between temporal fluctuations and variation within the population is similar to ergodicity in statistical physics — the time average of a system equals the ensemble average of identical systems.

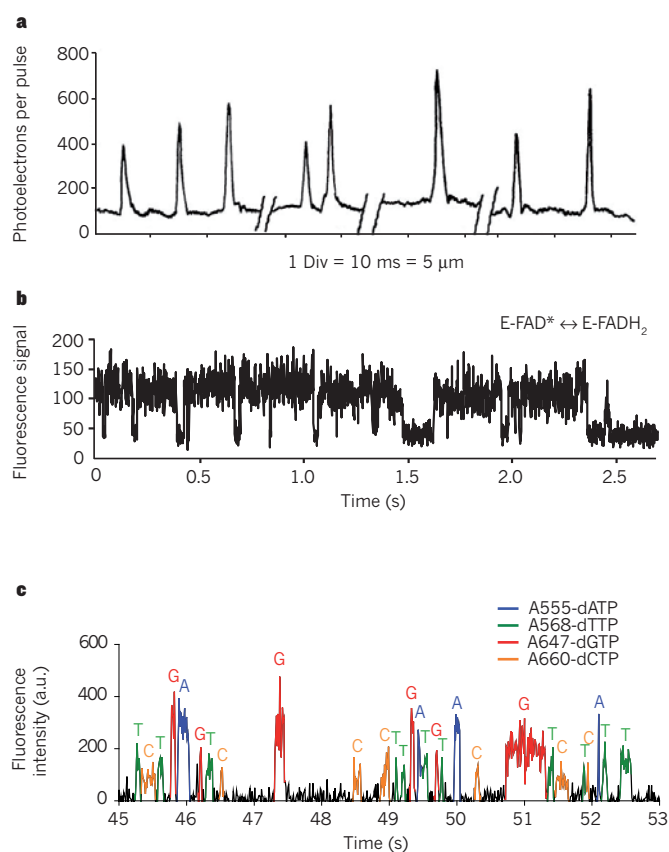
Here we review recent single-molecule experiments that provide quantitative descriptions of the central dogma in living bacterial cells, although the strategies and technical advances highlighted are applicable to future studies in eukaryotic cells. We show that single-molecule stochastic events have important biological consequences, such as determining the phenotype of a cell.

### Imaging single molecules in living cells

To image a particular biomolecule in a living cell with fluorescence microscopy, specific labelling is required. The advent of genetically encodable fluorescent proteins has provided the highest specificity so far, with minimal perturbation for live-cell imaging<sup>23</sup>, allowing real-time observations of fusion proteins of interest. Although the weak signal of a single fluorescent-protein molecule is detectable *in vitro* using a fluorescent microscope together with a combination of laser excitation and modern charge-coupled-device detectors (Fig. 3a), the detection of single fluorescent-protein reporters in living cells is challenging owing to strong cellular autofluorescence. This obstacle can be partly overcome by selecting fluorescent proteins that are spectrally separated from the autofluorescence, which is generally blue-green<sup>24</sup>. Yellow- or red-emitting fluorescent proteins are therefore favourable for live-cell single-molecule imaging. Furthermore, in the same spirit as Hirschfeld's experiment, the signal can be improved by reducing the detection volume to minimize autofluorescence background. For example, total internal reflection fluorescence microscopy (TIRFM) can limit the axial depth by illuminating with an evanescent wave that penetrates only a few hundred nanometres into a sample (Fig. 3b). TIRFM is therefore ideal for studying membrane protein dynamics<sup>25,26</sup>, but it does not allow imaging of the whole cell body. However, single fluorescent-protein imaging using wide-field illumination is possible in bacterial cells owing to their compact sizes (Fig. 3a).

In living eukaryotic cells, imaging a single fluorescent protein is more difficult. A typical mammalian nucleus is 5–10  $\mu\text{m}$  in diameter, compared with 1  $\mu\text{m}$  for a bacterial cell. In a wide-field microscope, such a large cell volume gives rise to a strong out-of-focus autofluorescence signal, which overwhelms the signal of a single fluorescent protein. Probing DNA–protein interactions therefore requires three-dimensional (3D) sectioning. Although confocal fluorescence microscopy with one-photon excitation could be used, it also causes photobleaching outside the focal plane<sup>27</sup>. One solution to this problem is to use two-photon fluorescence microscopy<sup>28,29</sup> (Fig. 3c), which allows localized excitation only at the laser focus, considerably reducing out-of-focus photobleaching while providing 3D sectioning in living eukaryotic cells. But, like confocal microscopy, it requires point scanning, thus limiting its time resolution. Alternatively, sheet illumination<sup>30,31</sup>, in which a thin light sheet illuminates only the image plane (Fig. 3d), provides low fluorescence background and high sensitivity, as well as high temporal resolution, because it does not require point scanning. These techniques are being adapted for single-fluorescent-protein imaging in living eukaryotic cells.

In a bacterial cell, a freely diffusing protein is difficult to image because its fast diffusion spreads the signal throughout the whole cell<sup>32,33</sup>. However, if a single fluorescent protein is localized, it can be imaged above the cellular autofluorescence background<sup>34</sup>. This method, termed detection by localization (Fig. 4a), works as long as there is only

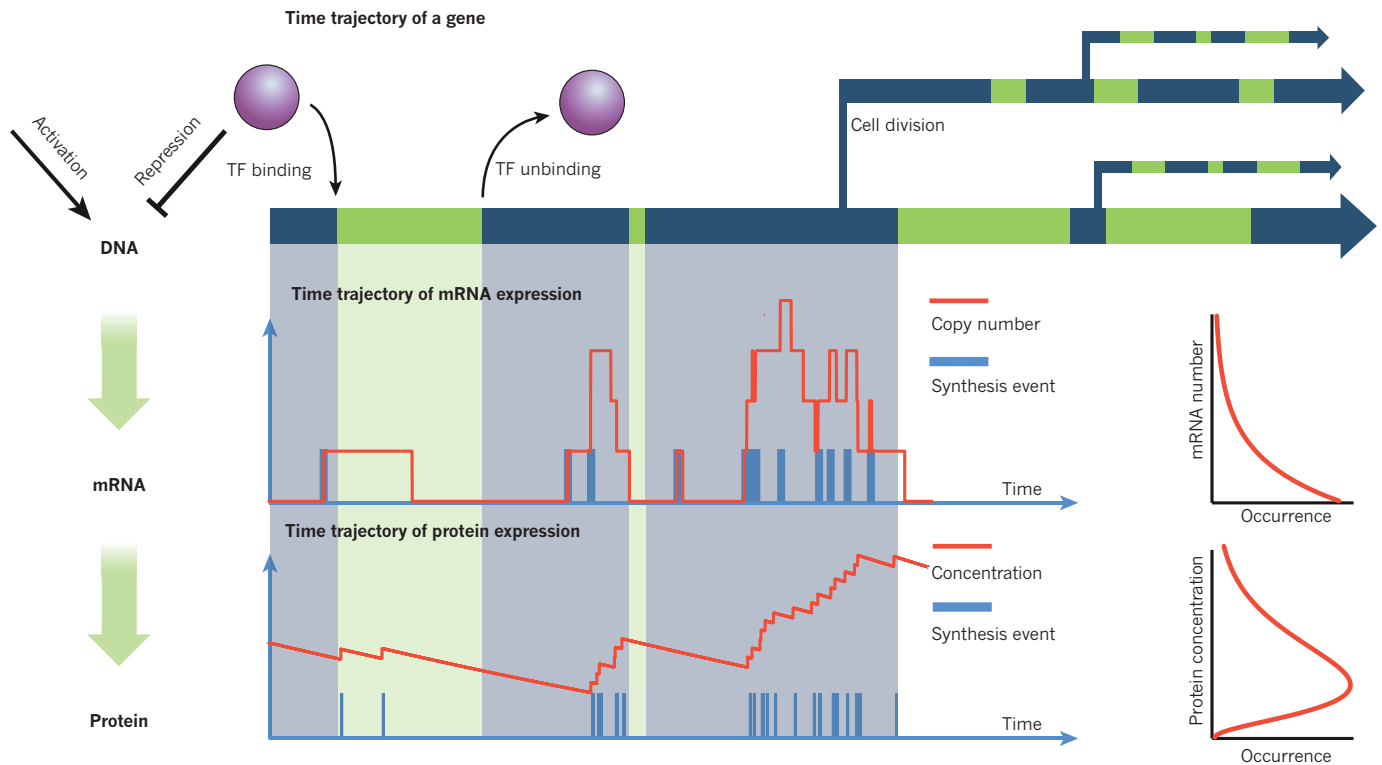


**Figure 1 | Stochastic nature of single-molecule processes.** **a**, Optical imaging of single protein molecules at room temperature. In his 1976 work, Hirschfeld demonstrated the detection of single protein molecules using a fluorescence microscope. A line scan of eight protein molecules was recorded. Div, division; adapted, with permission, from ref. 1. **b**, Stochastic turnovers of a single enzyme molecule. The fluorescence signal of a cholesterol oxidase molecule (E) shows stochastic switching between a fluorescent (oxidized flavin, FAD\*) and non-fluorescent (reduced flavin, FADH<sub>2</sub>) state as enzymatic turnovers take place. Adapted from ref. 9. **c**, Single-molecule DNA sequencing. A single DNA polymerase is used to sequence DNA by incorporating fluorescently labelled nucleotides of four different colours. Although each incorporation happens stochastically with variable waiting times, the overall time for DNA replication, which is a sum of many sequential steps, is narrowly distributed. a.u., arbitrary units; adapted, with permission, from ref. 11.

one immobilized molecule in a diffraction-limited volume (less than 10 molecules within a bacterial cell). Detection by localization can be done by tethering on a membrane<sup>34</sup>, or by specific or even transient nonspecific binding to DNA<sup>35</sup>.

In cases in which the frame rate of the camera is insufficient to detect transient localization (<10 ms), a shorter pulse of laser excitation can be provided with each imaging frame<sup>33,35</sup>, an idea borrowed from strobe photography. Detection by localization therefore allows single-molecule observations with millisecond time resolution.

The width of a single-molecule image is about half of the optical wavelength, owing to the diffraction limit. However, the accuracy of determining the centre position of a single isolated fluorescent protein can be as high as a few nanometres<sup>36</sup>. To image more concentrated samples, higher spatial resolution can also be achieved by selectively observing only one molecule at a time using photoactivatable fluorescent proteins. This is the idea behind recent developments in single-molecule-based super-resolution imaging, such as stochastic optical reconstruction microscopy<sup>37</sup> and photoactivated localization microscopy<sup>38,39</sup>, in which high-resolution images are reconstructed from many single-molecule images. Future applications of super-resolution techniques will probably change the way we view intracellular



**Figure 2 | Central dogma at the single-molecule level.** In a living bacterial cell, there is usually one copy of a particular gene, which is regulated by transcription factors (TFs), and transcribed into mRNA and translated into protein. A rate-limiting event, such as transcription-factor binding to and unbinding from DNA,

in this single-molecule process results in stochasticity. The expression levels of mRNA (middle panel) and protein (bottom panel) show temporal fluctuations in a single-cell lineage. This gives rise to variations of mRNA and protein copy numbers among a population of cells at a particular time (right panels).

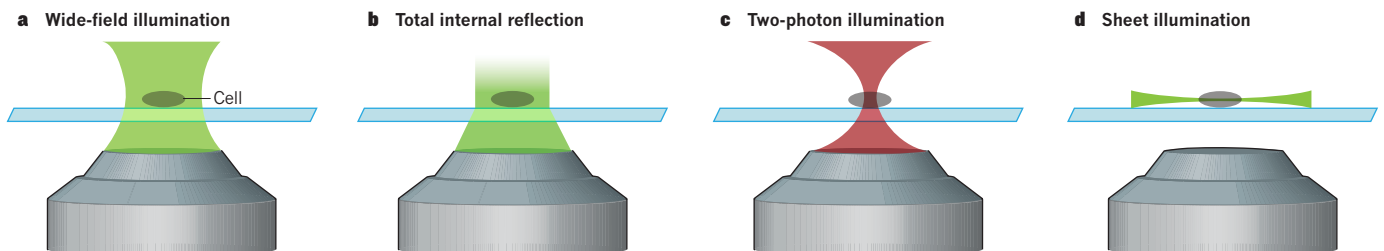
processes<sup>40</sup> such as gene expression. Single-fluorophore detection, as discussed above, remains a prerequisite for super-resolution imaging.

### Transcription-factor dynamics

As the first step of gene expression, transcription factors must bind to or unbind from DNA in response to environmental signals. Because transcription factors interact with DNA at one location, gene expression is stochastic when the binding and unbinding of a transcription factor become rate limiting (Fig. 2). In the classic example of the *lac* operon, the transcription factor known as the *lac* repressor (LacI), which is expressed at fewer than five copies per cell<sup>35</sup>, binds to or unbinds from operator sites to control transcription. With detection by localization, a single *lac* repressor fused to yellow fluorescent protein (YFP) can be visualized when bound to its operator in the *lac* operon<sup>35</sup>. When the inducer isopropyl- $\beta$ -D-thiogalactoside (IPTG) is added to the cell, localized fluorescent spots disappear as a result of LacI dissociation

(Fig. 4b). This living-cell assay allows single-molecule measurements of transcription-factor dissociation kinetics.

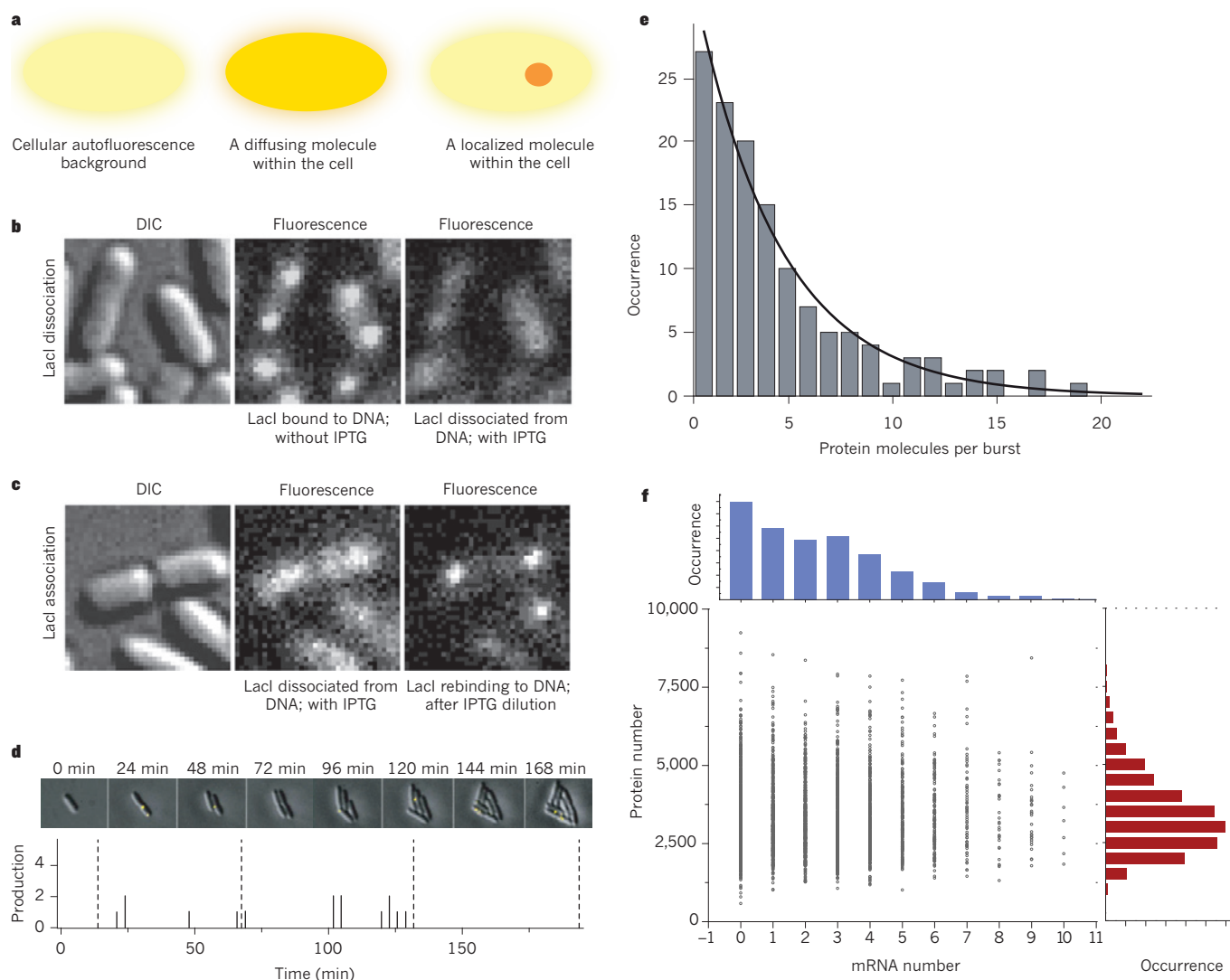
In addition, the binding kinetics can be measured. When IPTG is removed from the medium, the localized signal reappears, indicating the rebinding of LacI (Fig. 4c). This experiment allowed the first measurement of the time required for a LacI molecule to find a vacant operator site on DNA. It takes less than 360 s for one repressor to search for one specific binding site<sup>35</sup>. This 360-s search time is a result of complex molecular processes. The protein–DNA search problem was extensively studied in the 1970s and 1980s<sup>41,42</sup>. It was observed that the DNA-binding rate constant of transcription factors significantly exceeds that expected from the 3D diffusion limit for bimolecular binding<sup>43</sup>. This observation led to the prevailing model of facilitated diffusion. For a transcription factor or any DNA-binding protein to find a target sequence on DNA, it first binds somewhere along the DNA nonspecifically and undergoes 1D diffusion in search of the target. If the target is not found, the



**Figure 3 | Methods for imaging single molecules in living cells.** Single-molecule fluorescence can be imaged using multiple laser illumination geometries that reduce the probe volume. **a**, In wide-field illumination, the entire cell is subject to laser exposure. For bacterial cells that have small volume, no further probe volume reduction is necessary. **b**, With total internal reflection, only the region within a few hundred nanometres of the coverslip

is illuminated. This method is often used to image single membrane proteins, but cannot detect molecules deep in the cells. **c**, Two-photon excitation suppresses out-of-focus background, but suffers from slower time resolution owing to the need for point scanning. **d**, Sheet illumination has reduced background, as well as increased time resolution, because it does not require point scanning.





**Figure 4 | Real-time measurements of gene expression with single-molecule sensitivity.** **a**, Detection by localization. The cellular autofluorescence makes it difficult to detect a freely diffusing fluorescent protein. However, a localized single molecule can be imaged above the autofluorescence background. **b**, Detection of single transcription factors in living cells. A *lac* repressor (LacI) labelled with YFP can be imaged when bound to its operator site on DNA. The localized fluorescence disappears after dissociation caused by the inducer IPTG. DIC, differential interference contrast microscopy; adapted from ref. 35. **c**, Rebinding of LacI to the operator on dilution of IPTG, as evident from the reappearance of the fluorescence localization. The average rebinding time is 60 s, which can be explained by the facilitated diffusion model for a target search. Adapted from ref. 35. **d**, Real-time observation of protein synthesis under repressed

transcription factor dissociates from the DNA to avoid a long search time imposed by 1D diffusion. The 3D diffusion through the cytoplasm is much faster, allowing the transcription factor to reach distant segments of DNA quickly. This combined 1D and 3D search is repeated until the transcription factor finds the DNA segment containing the target sequence. With single-molecule experiments, one can probe these phenomena in real time and quantify the process.

In a series of single-molecule studies *in vitro*, 1D diffusion has been directly observed for fluorescently labelled transcription factors and other DNA-binding proteins along nonspecific DNA under a microscope<sup>44–49</sup>. The observed 1D diffusion rate (of the order of  $0.05 \mu\text{m}^2 \text{s}^{-1}$ ) is much slower than the 3D diffusion in a living cell ( $\sim 3 \mu\text{m}^2 \text{s}^{-1}$ ) because the 1D diffusion of the transcription factor is

coupled to simultaneous rotation around the DNA, tracking the pitch of the DNA double helix<sup>50,51</sup>. In the *in vitro* experiments, low salt concentrations were used to ensure long nonspecific residence times so that long trajectories of 1D diffusion could be recorded. In a living cell, high salt concentration shortens the residence time, but the diffusion constant often remains the same<sup>46</sup>. Consequently, the number of bases inspected in each 1D search segment is reduced.

A key question is whether such facilitated diffusion occurs in living cells. Recent single-molecule experiments suggest that it does. During the search process, a transcription factor spends 90% of its time on nonspecific DNA, and the residence time of nonspecific binding is less than 5 ms (ref. 35). Given the 1D diffusion constant *in vitro*, the protein inspects  $\sim 100$  base pairs (bp), which implies a 100-fold acceleration

of target search compared with the case with no 1D diffusion<sup>52</sup>. This observation is consistent with mounting evidence that the length of the DNA segment that a transcription factor inspects is shorter than 1,000 bp<sup>53–57</sup>, the value estimated from early *in vitro* experiments<sup>43</sup>. This 100-bp range indicates that for a  $5 \times 10^6$  bp genome, a transcription factor must inspect  $5 \times 10^6/100 = 5 \times 10^4$  segments before reaching the target site. Therefore, the total search time for one transcription factor in a cell is  $\sim 5 \times 10^4 \times 5 \text{ ms} = 250 \text{ s}$ , in close agreement with the measured search time<sup>35</sup>.

The combination of different single-molecule approaches has resolved the search problem and led to a quantitative understanding of the facilitated diffusion of transcription factors in bacteria. Similar single-molecule experiments should be able to address the same search problem in mammalian cells, which is complicated by nucleosomes.

## Gene expression in real time

Transcription-factor binding or unbinding leads to transcription and translation. Although the central dogma has been well established, real-time observation and quantitative description of transcription and translation in a single cell, at the single-molecule level, have only become possible in recent years. These studies have yielded unexpected observations of these fundamental processes in living cells<sup>34,58–62</sup>.

We first discuss protein production, as it is better understood at the single-molecule level under repressed (non-induced) conditions. Under these conditions, single-molecule experiments have shown that proteins are synthesized in bursts<sup>34</sup>, and that the characteristics of the bursts can be understood quantitatively at the molecular level. The stochastic production of individual molecules of a YFP-fused membrane protein has been monitored in real time in *Escherichia coli*<sup>34</sup> (Fig. 4d). Newly synthesized YFPs were visualized one by one as diffraction-limited spots through detection by localization, and they were purposely photobleached after being detected. A fast-maturing YFP, Venus<sup>63</sup>, was used to achieve 7-min time resolution in the observation of translation. Using this approach, translational bursting from the *lac* operon under repressed conditions was observed<sup>34</sup>. Each burst creates four proteins on average, at a frequency of about one burst per generation time (although not synchronized to the cell cycle). The number of bursts per cell cycle follows the Poisson distribution.

Because it was shown that each burst results from transcription of a single mRNA (generated owing to the occasional dissociation of the LacI repressor), the observed translational burst must therefore be due to several rounds of ribosomal initiation on the same transcript. This transcript is degraded by nucleases with a stochastic cellular lifetime that is exponentially distributed with a time constant of 1.5 min. The longer an mRNA lives, the more proteins it produces. Consequently, as theoretically predicted in the 1970s<sup>64–66</sup>, the burst size is exponentially distributed (Fig. 4e). This observation of exponentially distributed protein copy numbers per burst was independently confirmed by another single-molecule assay using  $\beta$ -galactosidase activity as a reporter<sup>62</sup>. As we discuss later, such stochastic expression due to transcription-factor unbinding can be important in determining how a gene is induced in the presence of external stimuli<sup>67</sup>.

Under repressed conditions in *E. coli*, the mRNA production is Poissonian. Under induced conditions, however, mRNA too is produced in bursts. One widely adopted method to detect single mRNA molecules in living cells uses the bacteriophage coat proteins (MS2) that stably bind to specific RNA sequences<sup>68</sup>. To visualize single copies of mRNA, cell lines are engineered to express both MS2–green fluorescent protein (GFP) and mRNA containing several MS2-binding sites. First developed by the Singer group, this method allows real-time observation of transcript production, and is ideal for probing transcriptional dynamics in living cells by tracking and counting single mRNA–MS2–GFP complexes<sup>69,70</sup>. A caveat is that the secondary structure associated with the binding sites and MS2 binding often interferes with the native mRNA degradation pathways<sup>71</sup>, preventing the profiling of endogenous mRNA expression levels.

When MS2-containing mRNA is expressed under fully induced

conditions, the production of transcripts is found to be intermittent<sup>58</sup>. If transcript production were to have a single rate-limiting step, such as RNA polymerase binding or initiation, the waiting time between the birth of each mRNA would be exponentially distributed, and the copy-number distribution would be Poissonian (with a variance equal to the mean). Surprisingly, short bursts (average 6 min) of mRNA synthesis followed by long periods (average 37 min) of inactivity have been observed<sup>58</sup>. The burst-like transcription is similar to that shown in Fig. 4d, even though there is no known transcription-factor binding or unbinding in this case. This burst-like transcription was also observed using fluorescence correlation spectroscopy on MS2-bound mRNA in *E. coli*<sup>59</sup>, as well as in eukaryotic cells<sup>60,61</sup>. Although the overall waiting time between each mRNA synthesis event is not exponentially distributed<sup>58</sup>, the waiting times for transition between the active and the inactive states are. Accordingly, the copy-number distribution is super-Poissonian, meaning that the variance of the distribution is greater than the mean. In other words, the cell-to-cell variation is significantly greater than what would be expected from a single rate-limiting process.

This important finding pointed out that transcription from a supposedly constitutive promoter is not as simple as RNA polymerases transcribing with a constant flux. Rather, it is a much noisier process, and the origin of this noise is unknown. Possible candidates include the role of nucleoid-associated proteins that are analogous to eukaryotic histones, global fluctuations of chromosome supercoiling states and RNA polymerase availability. *In vivo* single-molecule approaches are poised to further reveal the workings of these fundamental processes.

## Characterization of cell-to-cell variation

Under steady-state conditions, temporal fluctuations of gene expression in each cell lineage, as discussed in the previous section and Fig. 2, lead to variation in copy number in an isogenic population of cells. A typical copy-number distribution, which is often asymmetrical, is shown in Fig. 4f. A rigorous mathematical relationship between fluctuations in expression and the distribution of protein copy number in a population of cells has been lacking. A log-normal function has often been used as a convenient phenomenological fit, but it offers no physical insight.

For each gene, the dynamics of the central dogma can be described by two parameters — the burst frequency,  $a$ , which is the number of bursts per cell cycle; and the burst size,  $b$ , which is the average number of molecules produced per burst. Experimentally,  $a$  and  $b$  can be determined by single-cell trajectories, such as in Fig. 2. Alternatively, the fact that temporal fluctuations in a cell lineage are related to cell-to-cell variation of copy numbers suggests that  $a$  and  $b$  can also be inferred from a population of isogenic cells at a particular moment, as observed with a microscope or flow cytometer.

To find the relationship, we needed a governing equation for gene-expression dynamics. This is the chemical master equation, which was first used by Delbrück<sup>72</sup> in 1940. In the late 1970s, the chemical master equation was applied to obtain protein copy-number distributions resulting from stochastic gene expression<sup>64,65</sup>. It was not until a decade ago that this approach regained attention<sup>66,73,74</sup>. Given the chemical kinetics scheme and rate constants connecting all the macromolecules involved in the central dogma, one can, in principle, solve the chemical master equation, which naturally yields time-dependent fluctuations. In practice, this can be simulated numerically using the Gillespie algorithm<sup>75</sup>. Under certain conditions, analytical results can be obtained. For example, under steady-state conditions with uncorrelated and exponentially distributed bursts, the chemical master equation can be solved<sup>76</sup>, and the protein copy-number distribution,  $p(n)$ , can be approximated as a gamma distribution when the copy number ( $n$ ) is approximated as a continuous variable<sup>77</sup>:

$$p(n) = n^{a-1} e^{-n/b} / b^a \Gamma(a)$$

The gamma distribution has two kinetic parameters —  $a$  and  $b$ , as defined earlier — providing a clear physical interpretation of the



copy-number distribution. This mathematical relationship allows extraction of intrinsic kinetic parameters ( $a$  and  $b$ ) from fitting a gamma function to the measured copy-number distribution. At low expression levels, the values for  $a$  and  $b$  determined in this way are consistent with those derived from the single-cell trajectories<sup>34,62</sup>. As discussed in the next section, the cell-to-cell variation at high expression levels is more complicated but remains well described by a gamma distribution.

### Gene-expression profiling

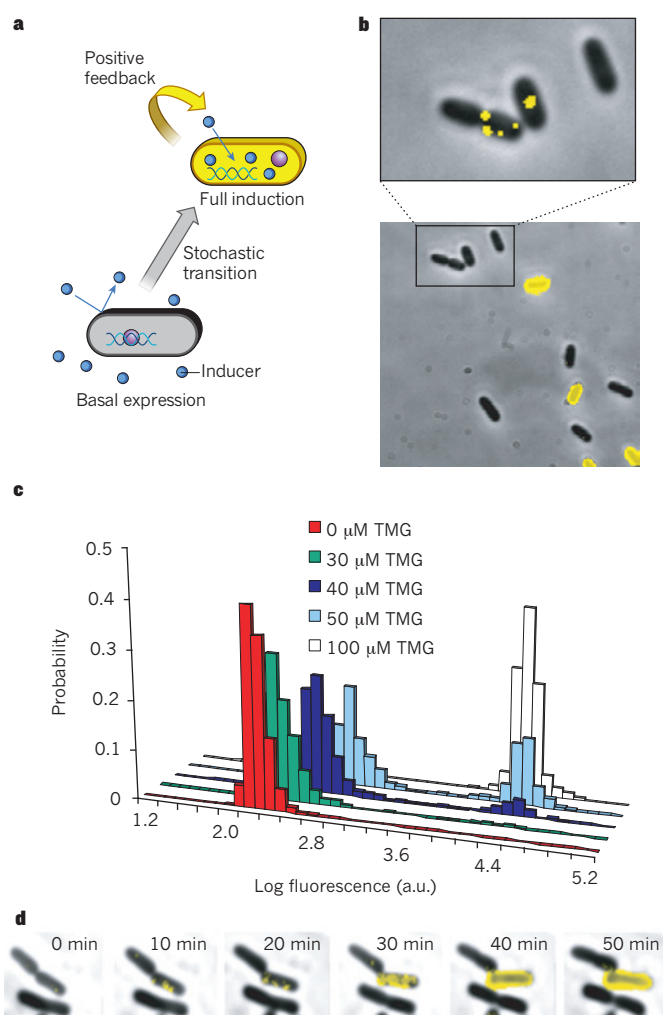
The ability to image single molecules in bacteria has offered an opportunity to profile protein expression globally, at any abundance level. Pioneering work using a yeast GFP fusion library<sup>78</sup> surveyed the cell-to-cell variation of more than 2,500 genes under various growth conditions, yielding several important observations<sup>79,80</sup>. First, the noise, or the variance divided by the mean squared, scales inversely with abundance. Second, the deviation of noise in a particular gene away from the global trend reflects the protein function and perhaps the underlying regulation. However, because single-molecule sensitivity in yeast cells had not been achieved at the time, 30% of the genes that were weakly expressed in the GFP library were undetectable.

To profile global variation at all expression levels, an *E. coli* YFP fusion library was constructed, and included more than 1,000 genes with expression levels ranging from 0.1 to  $10^4$  proteins per cell<sup>21</sup>. Of all the tagged proteins, approximately 99% of the copy-number distributions are well fit with the gamma distribution. About 50% of the proteins are expressed at an average level of fewer than ten molecules per cell, which argues for the necessity of single-molecule sensitivity in single-cell analyses.

Protein-expression noise has two distinct scaling properties relative to the mean. Below ten molecules per cell, the noise is inversely proportional to protein abundance. This scaling is the same as that observed in yeast, indicating that the noise from random birth and death of molecules, also known as intrinsic noise<sup>81,82</sup>, dominates the expression variation for low-abundance proteins. By contrast, at abundances above ten molecules per cell, the noise reaches a plateau of 30% and does not decrease any further. This noise plateau is common, or 'extrinsic', to most high-abundance proteins, as the expression levels of different proteins have a large covariance from cell to cell. Notably, time-lapse movies have shown that the extrinsic noise fluctuates at a timescale much longer than the cell cycle, suggesting that a slow global-regulation process is at work<sup>83</sup>.

At the transcriptional level, the same YFP library has been used to simultaneously survey mRNA and protein variation for 137 highly expressed genes<sup>21</sup>. Instead of labelling with MS2, which requires further cloning steps, mRNA was visualized using single-molecule fluorescence *in situ* hybridization (FISH)<sup>84,85</sup> in fixed cells. Unlike conventional approaches that use several hybridization probes against the mRNA, the YFP mRNA was targeted using a universal singly labelled FISH probe optimized for both hybridization efficiency towards its targets and specificity against off-targets. It was found that, even for highly expressed genes, the average mRNA copy number is fewer than five per cell. Among a population of genetically identical cells, every mRNA species has a distribution that is broader than a Poisson distribution (Fig. 4f), which is related to the transcriptional bursts observed in the real-time experiments and suggests that this is a general phenomenon for most genes.

The simultaneous profiling of mRNA and protein<sup>21</sup> revealed that the mRNA and protein copy numbers of a single cell for any given gene are uncorrelated; that is, a cell that has more mRNA molecules than average does not necessarily have more proteins (Fig. 4f). This perhaps counter-intuitive result can be explained by the fact that mRNA has a much shorter lifetime than protein in bacteria<sup>19</sup>. This finding argues for the necessity of single-cell proteomics analyses, and offers a warning for interpretations of single-cell transcriptome data, at least for bacteria. A mammalian cell, by contrast, has comparable mRNA and protein lifetimes, and hence is expected to have more-correlated mRNA and protein levels than a bacterial cell.



**Figure 5 | Phenotype switching due to a single-molecule event.** **a**, Bistability of the *lac* operon. The positive feedback by the normally repressed Lac permease (LacY, labelled with YFP) results in bimodal distribution at an intermediate inducer concentration, with two distinct phenotypes: strongly or weakly fluorescent. **b**, Fluorescence-microscope images show two phenotypes. The copy number of LacY in uninduced cells ranges from 0 to 10 molecules per cell, suggesting that one molecule of LacY is not enough to trigger the positive feedback. **c**, Bimodal distribution of LacY expression for isogenic cells at intermediate concentrations of an inducer (TMG, a lactose analogue). **d**, Time-lapse images capture the transition of a cell from one phenotype to another. A large expression burst of LacY (~300 molecules) is necessary to trigger the switching, which results from the complete dissociation of a single transcription factor, LacI, from DNA. This experiment shows that a low-probability, single-molecule stochastic event can determine cell fate. Adapted from ref. 67.

### Gene regulation and phenotypic switching

How cells with identical genomes have different phenotypes is an interesting question. Phenotypes are the physical, chemical and physiological states of the cell as related to function, determined by both the genome and environment. Given the ubiquitous and substantial noise described earlier, it is evident that the phenotype of a cell cannot be solely defined by its transcriptome and proteome. Cells can tolerate rather large noise in protein and mRNA abundance while tightly maintaining their phenotypes. A compelling question is what molecular actions dictate the transition between phenotypes.

In some cases, the cell phenotype can be clearly defined when there are bimodal or multimodal distributions of proteins, in contrast to the unimodal copy-number distribution that is most often observed<sup>21</sup>. As shown in Fig. 5b, a population of isogenic *E. coli* cells, in which

Lac permease is labelled with YFP, shows bistability. The *lac* operon in *E. coli*, consisting of *lacZ*, *lacY* and *lacA* genes, is normally repressed by the *lac* transcription-factor repressor (LacI) in the absence of an inducer (Fig. 5a). When the inducer is present, it inactivates LacI and triggers expression of the *lac* operon<sup>86</sup>. The synthesis of the permease increases the inducer influx, which inactivates more LacI, creating a positive feedback on permease expression<sup>87</sup>. Without an inducer, no cells are induced, whereas with high inducer concentrations, all cells are induced. At moderate inducer concentrations, only a fraction of the cells are induced (Fig. 5c). This bistability is controlled by the positive feedback of the *lac* operon<sup>88</sup>.

Bistability is commonly exploited by bacteria to generate alternative phenotypes<sup>89</sup>, such as persistence against antibiotics<sup>90</sup>, lysis or lysogeny after phage infection<sup>91</sup> and induction of the *lac* operon in *E. coli*<sup>87</sup>. Although much is known about the genetic switches, what drives the transition between two phenotypic states is unclear in many cases. How does a single cell make a decision about which phenotype to choose? With single-molecule imaging, uninduced *E. coli* cells have been shown to contain 0–10 copies of the permease enzyme, which is below the threshold for positive feedback (more than 300 molecules per cell)<sup>67</sup>. Transition to the fully induced state therefore requires a large burst of protein synthesis (Fig. 5d).

The transcription factor controlling permease synthesis, LacI, is a tetramer that binds to two DNA-binding sites, creating a DNA loop. Partial dissociation of LacI and rapid rebinding to DNA result in a single copy of mRNA and a small burst of permease, as was observed in the aforementioned real-time studies of the repressed *lac* promoters. When the repressor completely dissociates from both operators on DNA, a large burst of permease arises, because it takes a few minutes for the repressor to rebind<sup>35</sup>. Indeed, bistability was eliminated in strains without DNA looping<sup>67</sup>. It is the stochastic single-molecule event of complete repressor dissociation from DNA that triggers the cell's phenotypic switching.

## Looking forward

We have shown that in the case of the *lac* operon, the workings of the genetic switch can be quantitatively understood at the molecular level. This is an example of low probability, stochastic events of a single molecule having important biological consequences. Another simple example is point mutations in the course of evolution.

It is well recognized that such stochastic events are connected to cell-fate determination in other systems<sup>92</sup>. For example, there is considerable evidence that bacterial persistence against antibiotics is a stochastic process involving gene expression<sup>90</sup>. Persisters are not drug resistant but are drug tolerant. Drug resistance is related to a changing genome, whereas persisters have identical genomes, but different phenotypes. The phenomenon exists for many bacterial species and antibiotics. The molecular mechanism behind persistence is largely unknown, partly because the tools are not available. Understanding the molecular mechanism of persistence may be crucial to drug development, especially for diseases such as tuberculosis, caused by the bacterium *Mycobacterium tuberculosis*, which kills almost 2 million people every year worldwide. Single-cell gene-expression profiling may shed light on the mechanism of persistence.

Similarly, the reprogramming of somatic cells into induced pluripotent stem cells in the presence of certain transcription factors is also stochastic<sup>93</sup>. There are no elite cells, and every cell has a certain probability of being reprogrammed in the presence of some transcription factors, which is analogous to stochastic switching in the *E. coli lac* operon at low inducer concentrations. Yet, unlike the *lac* operon, the molecular mechanism is unknown. Extension of single-molecule approaches to mammalian cells and stem cells will allow real-time monitoring over long periods so that low-probability events with considerable biological consequences can be observed directly. We anticipate that the single-molecule approaches summarized in this Review will lead to more biological discoveries for many years to come. ■

- Hirschfeld, T. Optical microscopic observation of single small molecules. *Appl. Opt.* **15**, 2965–2966 (1976).  
**This paper presents the first single-molecule fluorescence imaging at room temperature, and the approach, which reduces the detection volume to enhance signal-to-background ratio, is widely used today in vitro and in living cells.**
- Shera, E. B., Seitzinger, N. K., Davis, L. M., Keller, R. A. & Soper, S. A. Detection of single fluorescent molecules. *Chem. Phys. Lett.* **174**, 553–557 (1990).
- Betzig, E. & Chichester, R. J. Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422–1425 (1993).
- Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
- Schmidt, T., Schutz, G. J., Baumgartner, W., Gruber, H. J. & Schindler, H. Characterization of photophysics and mobility of single molecules in a fluid lipid-membrane. *J. Phys. Chem.* **99**, 17662–17668 (1995).
- Xie, X. S. & Trautman, J. K. Optical studies of single molecules at room temperature. *Annu. Rev. Phys. Chem.* **49**, 441–480 (1998).
- Moerner, W. E. & Orrit, M. Illuminating single molecules in condensed matter. *Science* **283**, 1670–1676 (1999).
- Ha, T. et al. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl Acad. Sci. USA* **96**, 893–898 (1999).
- Lu, H. P., Xun, L. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282**, 1877–1882 (1998).  
**This article reports real-time observation and statistical analyses of stochastic enzymatic turnover and shows that on a single-molecule basis,  $k_{cat}$  is not a constant.**
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
- Eid, J. et al. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).  
**This paper reports real-time sequencing of DNA using a single DNA polymerase molecule, a compelling application of single-molecule enzymology in biotechnology.**
- Bouche, J. P. Physical map of a 470 x 10<sup>3</sup> base-pair region flanking the terminus of DNA replication in the *Escherichia coli* K12 genome. *J. Mol. Biol.* **154**, 1–20 (1982).
- Bremer, H. A stochastic process determines the time at which cell division begins in *Escherichia coli*. *J. Theor. Biol.* **118**, 351–365 (1986).
- Wang, M. D. et al. Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
- Blanchard, S. C., Kim, H. D., Gonzalez, R. L. Jr, Puglisi, J. D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc. Natl Acad. Sci. USA* **101**, 12893–12898 (2004).
- Kapanidis, A. N. et al. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
- Wen, J. D. et al. Following translation by single ribosomes one codon at a time. *Nature* **452**, 598–603 (2008).
- Cornish, P. V., Ermolenko, D. N., Noller, H. F. & Ha, T. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell* **30**, 578–588 (2008).
- Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S. & Cohen, S. N. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA* **99**, 9697–9702 (2002).
- Ishihama, Y. et al. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**, 102 (2008).
- Taniguchi, Y. et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).  
**This article profiles the transcriptome and proteome with single-molecule sensitivity in single cells, and shows that the protein copy-number distribution of most genes can be well fit with gamma distribution.**
- Xie, X. S., Choi, P. J., Li, G. W., Lee, N. K. & Lia, G. Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys.* **37**, 417–444 (2008).
- Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224 (2006).
- Andersson, H., Baechli, T., Hoechl, M. & Richter, C. Autofluorescence of living cells. *J. Microsc.* **191**, 1–7 (1998).
- Sako, Y., Minoguchi, S. & Yanagida, T. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nature Cell Biol.* **2**, 168–172 (2000).
- Iino, R., Koyama, I. & Kusumi, A. Single molecule imaging of green fluorescent proteins in living cells: E-cadherin forms oligomers on the free cell surface. *Biophys. J.* **80**, 2667–2677 (2001).
- Wilson, T. *Confocal Microscopy* (Academic, 1990).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
- Sanchez, E. J., Novotny, L., Holtom, G. R. & Xie, X. S. Room-temperature fluorescence imaging and spectroscopy of single molecules by two-photon excitation. *J. Phys. Chem. A* **101**, 7019–7023 (1997).
- Fuchs, E., Jaffe, J., Long, R. & Azam, F. Thin laser light sheet microscope for microbial oceanography. *Opt. Express* **10**, 145–154 (2002).
- Ritter, J. G., Veith, R., Veenendaal, A., Siebrasse, J. P. & Kubitschek, U. Light sheet microscopy for single molecule tracking in living tissue. *PLoS ONE* **5**, e11639 (2010).



32. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B. & Leibler, S. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* **181**, 197–203 (1999).
33. Xie, X. S., Yu, J. & Yang, W. Y. Living cells as test tubes. *Science* **312**, 228–230 (2006).
34. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).  
**This article monitors the real-time production of individual proteins by imaging newly synthesized YFP molecules, demonstrating translational bursts from a repressed promoter.**
35. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).  
**This article probes specific and nonspecific binding of a transcription factor to DNA, and measures the search time for a transcription factor to reach its target site in a living bacterial cell.**
36. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).
37. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods* **3**, 793–795 (2006).
38. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
39. Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
40. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
41. Richter, P. H. & Eigen, M. Diffusion controlled reaction rates in spheroidal geometry: application to repressor–operator association and membrane bound enzymes. *Biophys. Chem.* **2**, 255–263 (1974).
42. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
43. Riggs, A. D., Bourgeois, S. & Cohn, M. The *lac* repressor–operator interaction. 3. Kinetic studies. *J. Mol. Biol.* **53**, 401–417 (1970).
44. Kabata, H. *et al.* Visualization of single molecules of RNA polymerase sliding along DNA. *Science* **262**, 1561–1563 (1993).
45. Harada, Y. *et al.* Single-molecule imaging of RNA polymerase–DNA interactions in real time. *Biophys. J.* **76**, 709–715 (1999).
46. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl Acad. Sci. USA* **103**, 5752–5757 (2006).
47. Graneli, A., Yeykal, C. C., Robertson, R. B. & Greene, E. C. Long-distance lateral diffusion of human Rad51 on double-stranded DNA. *Proc. Natl Acad. Sci. USA* **103**, 1221–1226 (2006).
48. Tafvizi, A., Huang, F., Fersht, A. R., Mirny, L. A. & van Oijen, A. M. Tumor suppressor p53 slides on DNA with low friction and high stability. *Biophys. J.* **95**, L01–L03 (2008).
49. Kim, J. H. & Larson, R. G. Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic Acids Res.* **35**, 3848–3858 (2007).
50. Schurr, J. M. The one-dimensional diffusion coefficient of proteins absorbed on DNA. Hydrodynamic considerations. *Biophys. Chem.* **9**, 413–414 (1979).
51. Blainey, P. C. *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nature Struct. Mol. Biol.* **16**, 1224–1229 (2009).
52. Li, G. W., Berg, O. G. & Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nature Phys.* **5**, 294–297 (2009).
53. Stanford, N. P., Szczelkun, M. D., Marko, J. F. & Halford, S. E. One- and three-dimensional pathways for proteins to reach specific DNA sites. *EMBO J.* **19**, 6546–6557 (2000).
54. Winter, R. B., Berg, O. G. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry* **20**, 6961–6977 (1981).
55. Gowers, D. M. & Halford, S. E. Protein motion from non-specific to specific DNA by three-dimensional routes aided by supercoiling. *EMBO J.* **22**, 1410–1418 (2003).
56. Bonnet, I. *et al.* Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucleic Acids Res.* **36**, 4118–4127 (2008).
57. Porecha, R. H. & Stivers, J. T. Uracil DNA glycosylase uses DNA hopping and short-range sliding to trap extrahelical uracils. *Proc. Natl Acad. Sci. USA* **105**, 10791–10796 (2008).
58. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).  
**This article uses MS2–GFP to follow the production of individual mRNA molecules in real time in bacterial cells.**
59. Le, T. T. *et al.* Real-time RNA profiling within a single bacterium. *Proc. Natl Acad. Sci. USA* **102**, 9160–9164 (2005).
60. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr. Biol.* **16**, 1018–1025 (2006).
61. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
62. Cai, L., Friedman, M. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
63. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnol.* **20**, 87–90 (2002).
64. Berg, O. G. A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* **71**, 587–603 (1978).
65. Rigney, D. R. Stochastic-model of constitutive protein levels in growing and dividing bacterial cells. *J. Theor. Biol.* **76**, 453–480 (1979).
66. McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
67. Choi, P. J., Cai, L., Frieda, K. & Xie, X. S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446 (2008).  
**This article investigates the triggering mechanism of the *lac* operon bistable switch and shows that the dissociation of a single *lac* repressor is responsible for the phenotype switch to the induced state.**
68. Tyagi, S. Imaging intracellular RNA distribution and dynamics in living cells. *Nature Methods* **6**, 331–338 (2009).
69. Bertrand, E. *et al.* Localization of *ASH1* mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).  
**This study develops an approach to image individual mRNA molecules in living cells, using fluorescent protein-fused MS2 proteins that bind to an engineered mRNA sequence.**
70. Beach, D. L., Salmon, E. D. & Bloom, K. Localization and anchoring of mRNA in budding yeast. *Curr. Biol.* **9**, 569–578 (1999).
71. Johansson, H. E. *et al.* A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl Acad. Sci. USA* **95**, 9244–9249 (1998).
72. Delbrück, M. Statistical fluctuations in autocatalytic reactions. *J. Chem. Phys.* **8**, 120–140 (1940).
73. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).
74. Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc. Natl Acad. Sci. USA* **105**, 17256–17261 (2008).
75. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
76. Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys. Rev. Lett.* **84**, 5447–5450 (2000).
77. Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).  
**This article derives gamma distribution of protein copy numbers in a model with burst-like gene expression; the distribution is in agreement with experimental observations in reference 62.**
78. Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
79. Newman, J. R. *et al.* Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
80. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38**, 636–643 (2006).
81. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
82. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
83. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
84. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts *in situ*. *Science* **280**, 585–590 (1998).
85. Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science* **317**, 526–529 (2007).  
**This article used single-molecule FISH to show that the noise in the expression of a transcription factor determines whether a *Bacillus subtilis* cell becomes competent.**
86. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
87. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740 (2004).
88. Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553–566 (1957).
89. Dubnau, D. & Losick, R. Bistability in bacteria. *Mol. Microbiol.* **61**, 564–572 (2006).
90. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
91. Ptashne, M. *A Genetic Switch* 3rd edn (Cold Spring Harbor Laboratory Press, 2004).
92. Losick, R. & Desplan, C. Stochasticity and cell fate. *Science* **320**, 65–68 (2008).
93. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009).

**Acknowledgements** We thank L. Cai, J. Xiao, J. Yu, N. Friedman, J. Elf, P. Choi, H. Chen and Y. Taniguchi for their key contributions to the work described, and W. Greenleaf, P. Sims and H. Ge for their helpful comments on this review. This work is supported by the National Institutes of Health Director's Pioneer Award and the Bill & Melinda Gates Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to X.S.X. ([xie@chemistry.harvard.edu](mailto:xie@chemistry.harvard.edu)).

# Dynamic molecular processes mediate cellular mechanotransduction

Brenton D. Hoffman<sup>1</sup>, Carsten Grashoff<sup>1</sup> & Martin A. Schwartz<sup>1,2,3</sup>

**Cellular responses to mechanical forces are crucial in embryonic development and adult physiology, and are involved in numerous diseases, including atherosclerosis, hypertension, osteoporosis, muscular dystrophy, myopathies and cancer. These responses are mediated by load-bearing subcellular structures, such as the plasma membrane, cell-adhesion complexes and the cytoskeleton. Recent work has demonstrated that these structures are dynamic, undergoing assembly, disassembly and movement, even when ostensibly stable. An emerging insight is that transduction of forces into biochemical signals occurs within the context of these processes. This framework helps to explain how forces of varying strengths or dynamic characteristics regulate distinct signalling pathways.**

There is growing recognition that mechanical factors, such as applied forces or the rigidity of the extracellular matrix (ECM), crucially influence the form and function of cells and organisms<sup>1–3</sup>. Biological regulation has classically been understood through the concepts of solution chemistry, in which enzyme activities, reaction rates and affinities govern cellular processes. However, mechanotransduction, the conversion of mechanical forces into biochemically relevant information, contributes to numerous developmental, physiological and pathological processes and is a rapidly advancing area of current research<sup>1</sup>.

In the vasculature, blood flow exerts fluid shear stresses on the endothelial cells lining the vessels, whereas blood pressure stretches the vessel wall<sup>4</sup>. Shear stress is crucial for remodelling the primitive vascular plexus into a hierarchical vascular tree<sup>5</sup> and for patterning the cardiac outflow tract in developing mouse embryos<sup>6</sup>. Hypertension causes thickening of the arterial walls and is a major risk factor for cardiovascular diseases<sup>4</sup>. Atherosclerosis, the chronic cholesterol-dependent inflammation of artery walls, occurs preferentially at regions of disturbed flow such as branch points and areas of high curvature, where both the magnitude and temporal characteristics of the flow are disturbed. The development and pathobiology of bone and muscle also strongly depend on mechanical forces from weight and muscle contraction, whereas lung physiology and pathology are strongly influenced by forces from inflation<sup>1</sup>.

Tissue rigidity or stiffness affects many biological processes<sup>7</sup>. Tumours have long been identified by palpation, owing to local increases in tissue stiffness. Recently, these changes in the mechanical environment have been shown to be causal for tumour progression<sup>8</sup>. Fibrotic lung disease begins with a small change in tissue stiffness, which is sensed by cells, inducing more severe, irreversible remodelling<sup>9</sup>. Furthermore, the rigidity of the extracellular environment potentially controls the differentiation of mesenchymal stem cells<sup>10</sup> and the self-renewal of haematopoietic stem cells<sup>11</sup>. Developing scaffolds with tunable mechanical properties to control cellular behaviour has become a major effort in tissue engineering<sup>12</sup>.

Although applying forces to cells and altering the rigidity of their environment are clearly distinct processes, the underlying mechanisms of mechanotransduction seem to be similar<sup>2</sup>. A key event in rigidity sensing is the modulation of cellular contractility.

Cells on soft materials exert lower forces than cells on stiff materials, decreasing tension on force-bearing elements. These elements are the same whether forces are generated internally or externally<sup>7</sup>; thus, many of the cellular responses to distinct mechanical stimuli are similar. Another unifying principle is that the structures that generate and bear cellular forces are involved in sensing forces<sup>1</sup>. Therefore, cytoskeletal proteins such as actin and tubulin are crucial for mediating mechanical effects in nearly all systems<sup>2,13</sup> (Fig. 1a, b). Cellular adhesions, both to the ECM and to other cells, are also important, as they mechanically connect cells to their surroundings. Correspondingly, many of the candidate genes associated with diseases that can be considered 'mechanotransduction disorders' — such as aortic aneurism, heart failure, hypertension and muscular dystrophy — encode proteins involved in adhesion complexes, the cytoskeleton and the ECM<sup>14–17</sup>. There are often drastic changes in the protein composition, dynamics and mechanics of these structures during metastatic progression<sup>18</sup> and stem-cell differentiation<sup>7</sup>.

Although much progress has been made towards understanding mechanotransduction, a complete picture is lacking. Mechanotransduction is typically depicted as a series of rapid switch-like events, activated in response to step-like applications of force, which eventually lead to cellular responses. This level of detail, however, is insufficient to explain the cellular responses to dynamic mechanical stimuli often found in physiological settings.

In this Review, we first outline the basic features of the switch-like model of mechanotransduction. We divide this process into mechanotransmission, mechanosensing and mechanoresponse, and then highlight the limitations of the model. We also describe recent advances in our understanding of the dynamic processes regulating load-bearing subcellular structures and the behaviour of single molecules in response to applied forces. With guidance from mathematical models of adhesion assembly, these examples are used to develop a more complete model of mechanotransduction based on the concept that forces alter the rates of key subcellular processes to affect cell function. This perspective allows us to understand how cells respond to time-varying mechanical stimuli. We end by suggesting that the cell may function mechanically as a multiband pass filter in which stimuli with different temporal characteristics activate distinct signalling pathways that affect cell state and disease progression.

<sup>1</sup>Robert M. Berne Cardiovascular Research Center, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>2</sup>Department of Microbiology, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>3</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia 22908, USA.



## Switch-like models of mechanotransduction

Descriptions of mechanotransduction typically begin with the forces acting on cellular elements and end with the integrated response from the cell or tissue. These can be divided into three steps.

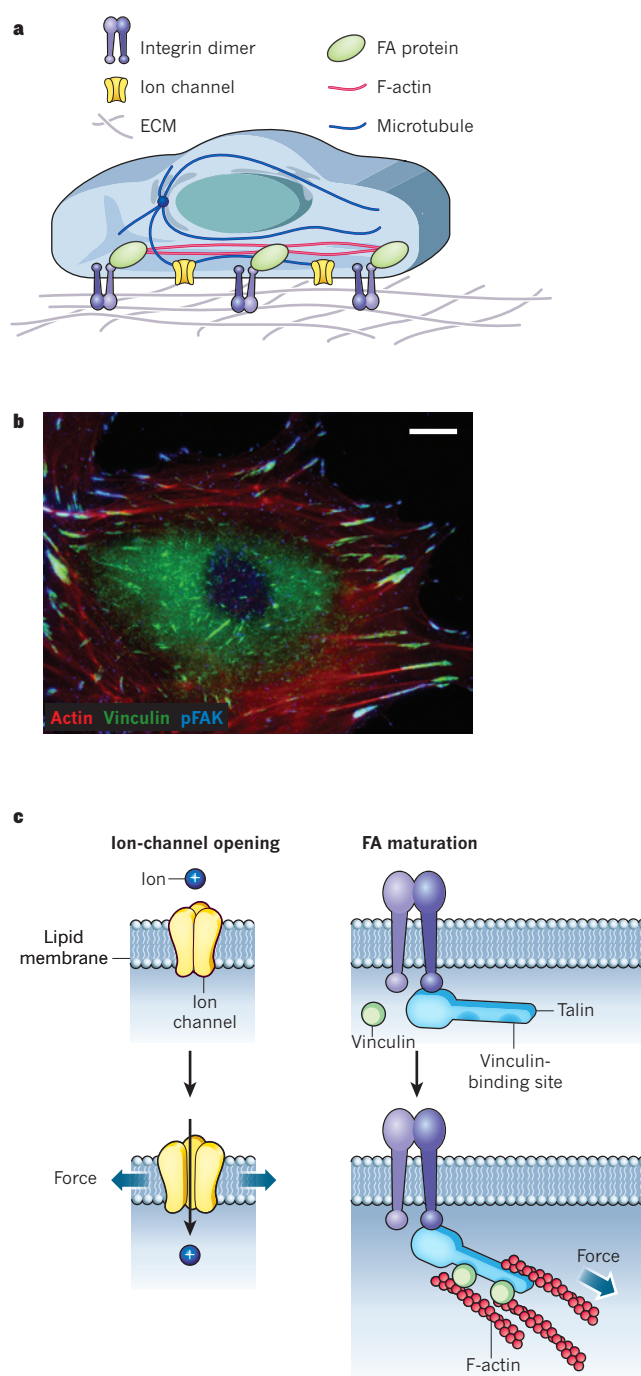
### Mechanotransmission

A force must be transmitted to mechanosensitive elements before it can be sensed. For example, the adhesion receptors that mediate cell–cell and cell–ECM contacts are strongly implicated in mechanotransduction<sup>19</sup>. Equally important are the cytoskeletal structures that adhesion receptors universally connect to, which allow adhesions to resist deformation from applied forces. The cytoskeleton is composed of filaments, such as F-actin, intermediate filaments and microtubules, that are relatively stiff on the micrometre-dimensional scale and are stable on minute-to-hour timescales<sup>20</sup>. This mechanical continuity allows forces to propagate relatively long distances along filaments in the cell, a process known as mechanotransmission. Fluid shear stress, for instance, is exerted on the apical domain of endothelial cells; yet the displacements in vimentin filaments are largest at select areas, often near cell–cell and cell–ECM adhesions<sup>21</sup>. This result correlates with data implicating junctional proteins in responses to shear<sup>22</sup>. Twisting of magnetic beads bound to integrins also showed long-distance effects, which were primarily transmitted by F-actin<sup>23</sup>, although some studies have proposed a role for microtubules<sup>24</sup> and intermediate filaments<sup>25</sup>. Using a laser trap to apply forces as small as 5.5 pN to actin stress fibres triggered an influx of calcium ions, presumably owing to the activation of mechanosensitive ion channels in the plasma membrane<sup>26</sup>. Cellular responses to force can also be extremely fast, of the order of hundreds of milliseconds<sup>24,27</sup>, consistent with direct mechanical effects.

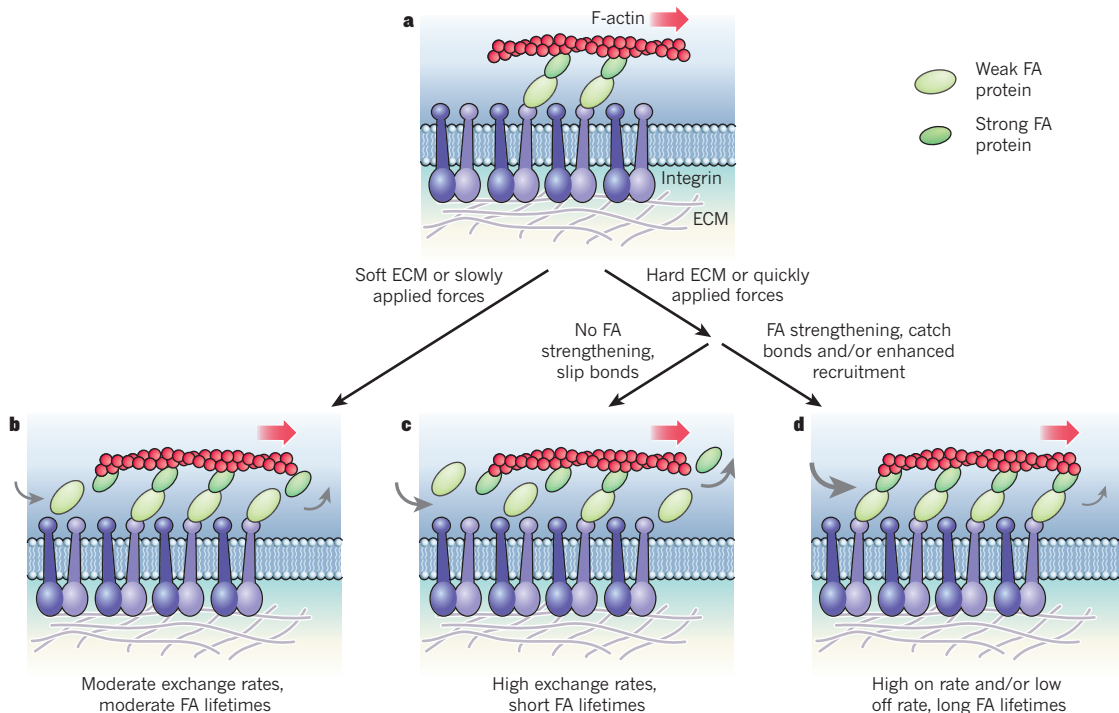
### Mechanosensing

Transmitted forces ultimately impinge on mechanosensitive macromolecules to alter their conformation and hence their function. Although the biological consequences of such events are specific to each system, the underlying physical response is similar; forces promote changes in protein conformation that accommodate the applied force. The best studied examples at the structural level are bacterial mechanically gated ion channels<sup>28</sup>, which open in response to increased lateral tension in the plasma membrane during osmotic swelling. Similar ion channels are present in all organisms and are essential for survival under changing osmotic conditions (Fig. 1c)<sup>29</sup>.

There is also evidence that the unfolding of protein domains under tension mediates responses to applied forces. The first reported instance was fibronectin, which self-assembles into fibrils in the ECM. The formation of fibronectin fibrils requires cell-generated force<sup>30</sup>. Conversely, purified fibronectin undergoes self-association when stretched *in vitro*<sup>31,32</sup>; fibril assembly is mediated by the unfolding of domains revealing cryptic-binding sites. Another well-studied example is talin-1, which connects integrins to F-actin, thereby transmitting forces between actomyosin filaments and the ECM<sup>33</sup>. Talin-1 binds to vinculin, which also links to F-actin and is recruited to adhesions in response to applied forces. Curiously, many of the vinculin-binding sites on talin reside within the interior of bundles composed of four or five  $\alpha$ -helices and are therefore inaccessible<sup>33</sup>. Both biochemical and cellular studies provide evidence that tension unfolds these bundles to expose vinculin-binding sites, thereby allowing vinculin recruitment<sup>34,35</sup> (Fig. 1c). Another protein in the cytoplasmic region of integrin-mediated adhesions is the adaptor protein p130<sup>Cas</sup> (also known as BCAR1). When phosphorylated on tyrosine residues by Src family kinases, p130<sup>Cas</sup> binds several guanine-nucleotide exchange factors (GEFs) that activate small GTPases<sup>36</sup>. Stretching cells enhances the phosphorylation of these tyrosines, leading to GEF binding and activation of Ras-related protein 1, widely known as Rap1 (refs 37–40). Studies with purified proteins have shown that stretching increases the susceptibility of p130<sup>Cas</sup> to phosphorylation, without changing the intrinsic activity of Src family kinases<sup>40</sup>. Although it is unclear how forces might be transmitted across



**Figure 1 | Switch-like models of mechanotransduction.** **a**, Cells are mechanically integrated structures, in which the ECM and actin cytoskeleton are connected by integrins and focal adhesion (FA) proteins. Microtubules and many ion channels are also integrated with this network. Forces can be applied directly through the ECM or transmitted through the cytoskeleton to mechanosensitive components, such as FAs, to mediate cellular response to forces. **b**, Immunostaining of a vascular smooth muscle cell. F-actin filaments (red) link to variably sized, punctate FAs, as shown by vinculin (green) and phosphorylated focal adhesion kinase (pFAK; blue) staining. The variable amount of pFAK staining in the FAs is indicative of different local signalling environments that are probably linked to distinct mechanical signals. Scale bar, 10 μm. **c**, A common mechanism of mechanotransduction is force-induced conformational change. For example, membrane tension can cause ion-channel opening. Also, talin connects the integrin cytoplasmic tail to F-actin; tension on talin exposes cryptic vinculin-binding sites, and the subsequent binding of vinculin (green) reinforces the linkage.



**Figure 2 | The focal-adhesion clutch.** **a**, Owing to forces from actin polymerization and myosin-dependent contractility, actin filaments flow backwards over FAs towards the nucleus. Through FA proteins that link actin to integrins, force is applied to the ECM. Force-dependent changes in FA exchange rates (arrows; see Box 1 for details) alter the dynamics and size of FAs. **b**, On soft surfaces or when external force is applied slowly, on rates and off rates are moderately high, and FAs have moderate lifetimes. **c, d**, By contrast, on stiff surfaces or when external forces are applied quickly,

distinct behaviours can be observed. In the absence of FA strengthening (**c**), molecular-linker dissociation rates increase (slip-bond behaviour). These proteins are sometimes replaced through rebinding, resulting in large exchange rates and short-lived FAs. FA strengthening is, however, associated with catch bonds, which slow FA–protein dissociation under force (**d**). This exposes cryptic binding sites and recruits proteins that reinforce the adhesion, and causes conformational changes in FA proteins that activate signalling pathways to recruit other molecular linkers, resulting in large, long-lived FAs.

p130<sup>Cas</sup> *in vivo*, these studies illustrate that forces can affect substrate availability through effects on protein conformation.

For physiologically significant mechanosensing, these initial conformational changes must be followed by a second step in which the new conformation triggers downstream events. This step can be fairly direct, as for the mechanosensitive ion channels discussed earlier. In other instances, changes in protein conformation, especially the opening of domains that contain cryptic sites, lead to the binding of proteins that mediate downstream events. This can result in reinforcement of the linkage, as in the case of talin and vinculin<sup>34</sup>, or the recruitment and activation of signalling proteins, as in the case of p130<sup>Cas</sup> (ref. 40). This general model is applicable to a wide range of mechanotransduction events in many systems<sup>1</sup>. Understanding in detail how protein domains change conformation under force and how subsequent events transpire has been the major direction in this field.

### Mechanoresponse

Ultimately, sensed mechanical signals influence information processing through complex cellular signalling and transcriptional networks that are not specifically force dependent. In many cases, these responses feed back to alter the mechanosensitive structures that initiated the responses. Both integrin-mediated and cadherin-mediated adhesions enlarge and strengthen in response to tension<sup>19</sup>. Distinct from the very rapid, direct recruitment described earlier, signalling pathways that are activated over minutes (such as the small GTPase RhoA, which stimulates the formation of actin stress fibres<sup>41</sup>) and gene-expression pathways that operate over hours or days (such as the induction of vinculin through serum response factor<sup>42,43</sup>) change the composition and structure of adhesions and the cytoskeleton.

Similar principles apply at the tissue level. High blood pressure, for instance, results in the thickening of artery walls to bear the increased

tension<sup>4</sup> and in hypertrophy of the left ventricle of the heart to allow stronger pumping against high back pressure<sup>44</sup>. Analogously, bone deposition increases under weight-bearing exercise<sup>1</sup>. These integrated responses depend on the intensity and time course of stimulation in ways that differ from the initial responses. For example, a single, brief interval of high blood pressure during exercise will stretch vessel walls and tax the heart but does not trigger compensatory arterial and cardiac remodelling, unlike sustained hypertension<sup>44</sup>.

### Limitations of switch-like models

In switch-like models of mechanotransduction, applied forces are instantaneously transmitted to load-bearing subcellular structures and induce conformational changes in mechanosensitive proteins. Different forces are sensed largely by conformational changes in protein domains that are stronger or weaker, and thus respond to forces of different magnitudes<sup>45</sup>. This view, however, seems to be incomplete. For example, the frequency of applied cyclic stretch or compression can have major effects. Steady stretch and cyclic stretch of equivalent magnitude induce distinct genes in endothelial cells<sup>46</sup> and induce differential phosphorylation of sites on focal adhesion kinase in rabbit aortas stretched *ex vivo*<sup>47</sup>. The frequency of applied cyclic stretch also determines endothelial alignment<sup>48</sup>. In aortic smooth muscle cells, stimulation of integrin activation and subsequent cellular alignment by cyclic stretch depends strongly on stretch duration and frequency<sup>49</sup>.

These observations are particularly relevant in the vascular system, in which shear stress and circumferential stretch in arteries undergo strong, time-dependent variations during the cardiac cycle. Furthermore, these variations are different in various parts of the vasculature<sup>50</sup> and correlate with the location of atherosclerotic lesion formation<sup>4</sup>. Notably, recent evidence demonstrates that particular

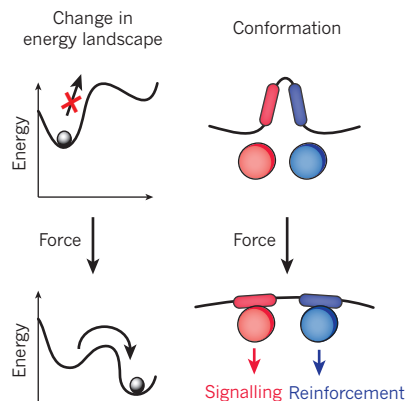


## BOX 1

# Physics of force-activated bond dissociation

The non-covalent bonds that mediate protein-protein interactions have finite lifetimes, ranging from milliseconds to days<sup>63</sup>. Applied force typically shortens the lifetime of a bond (like applying force to remove tape). In a molecular context, these are referred to as slip bonds<sup>63,64</sup>. There are also molecules in which the bond lifetimes increase, although not infinitely so, in response to applied forces<sup>64</sup>. These are called catch bonds and, although relatively rare, are often found in cytoskeletal and cellular adhesion structures.

Protein conformational changes can be understood as a type of slip bond in which the dissociation is internal, owing to non-covalent bonds between amino acids in a single protein instead of between two proteins<sup>63,65</sup>. Moreover, both processes can be represented in terms of an energy landscape in which two energy minima are separated by a high energy state that slows the rate of transition. Applied force acts catalytically to accelerate rates by lowering the energy requirement for the transition and by changing the free energy of the states, typically stabilizing



the more open conformation or unbound conformations. Thus, the likelihood of a conformational change depends on both the magnitude of the force and its duration<sup>63</sup>. These conformational changes mediate subsequent mechanotransduction events by exposing binding sites for signalling and cytoskeletal proteins (see Figure; boxes and circles denote binding sites and proteins, respectively; red denotes signalling proteins, and blue represents cytoskeletal proteins).

Furthermore, the bonds mediating protein conformations and protein-protein interactions tend to have similar affinities and force sensitivities. As protein dissociation will terminate the tension that induces conformational changes, these processes will in effect compete. This competition was recently studied using a single-molecule system that contained both a dissociable bond and a protein that could undergo force-dependent conformational opening<sup>66</sup>. Protein dissociation and conformational changes were both observed, but, interestingly, the frequency of conformational changes was enhanced at higher loading rates.

frequencies of mechanical stimulation preferentially activate the inflammatory pathways implicated in atherosclerosis, even when the total power applied to the cell is conserved (R. E. Feaver, B. D. Gelfand and B. R. Blackman, manuscript submitted). The characteristic time between the peaks in wall stretch and fluid shear stress may also crucially regulate endothelial activation<sup>51</sup>. These characteristics of mechanotransduction are not readily captured by switch-like models.

## Subcellular structures are dynamic

These time-dependent aspects of mechanotransduction can be attributed to the highly dynamic characteristics of the cellular components that bear and respond to force. Cell adhesions to the ECM go through a complicated, force-sensitive maturation process<sup>41</sup>. Nascent cell-ECM adhesions are very small structures (<1  $\mu\text{m}$  in diameter) at the edges of lamellipodia and usually disassemble within tens of seconds, or else mature into slightly larger focal complexes that persist for only a few minutes<sup>41</sup>. A fraction of these focal complexes mature into larger focal adhesions (FAs) that persist for tens of minutes. However, even within stable FAs, proteins are constantly exchanged, with lifetimes from tens of seconds to at most a few minutes<sup>52</sup>. Thus, even stationary FAs undergo rapid internal dynamics.

Detailed analyses from high-resolution techniques such as fluorescence speckle and correlation microscopy<sup>53,54</sup> have shown complex interactions between cytoskeletal and FA dynamics. Actin filaments constantly polymerize at the leading edge of the cell and flow backwards over the FAs, with the speed of this flow influenced by the nature of the linkages between the cytoskeleton, the integrins and the ECM. Actin flow is faster over areas with few FAs or in which the FAs undergo treadmilling towards the centre of the cell ('sliding FAs'), and slower in areas with stable FAs<sup>53</sup>. These and other results<sup>55</sup> led to the notion of a 'clutch' that controls force transmission between the flowing actin and the integrins (Fig. 2). Furthermore, in areas with stable FAs, integrins are immobile, and actin flows at 0.1–0.2  $\mu\text{m min}^{-1}$ ; however, different FA proteins have different velocities between these limits<sup>53</sup>. These results indicate the presence of many proteins that act as clutches, or force-sensitive linkages, within FAs.

Cadherin-dependent adhesions have not been studied in as much detail, but available evidence indicates similarities to FAs<sup>2</sup>. Several studies have shown that cell-cell contacts bear considerable forces<sup>56,57</sup> and, like FAs, they undergo dynamic, myosin-dependent elongation<sup>58</sup>. Applied forces<sup>56</sup> and stiffer substrata<sup>59</sup> enhance cell-cell contact assembly, indicating that these adhesions also undergo force-dependent adhesion strengthening. There is also evidence for actin flow along cell-cell contacts<sup>60</sup>. Although the other molecular players are different, vinculin is recruited to both structures in a myosin-dependent manner, thereby contributing to adhesion strengthening<sup>61,62</sup>. These data indicate that the dynamic properties of cell-cell contacts are regulated by processes physically similar to FA regulation.

## Single-molecule responses to dynamic forces

Studies of single molecules or pairs of molecules under applied force have rarely shown simple, switch-like behaviours. Of particular relevance, recent work on the effects of force application on the rates of bond dissociation shows that bonds mediating protein-protein interactions can either decrease or increase their average lifetime in response to applied force, referred to as slip or catch bonds, respectively<sup>63,64</sup> (Box 1). Notably, both the strength of the force and the rate of application affect the rates of protein conformational changes<sup>63,65</sup>. The development of assays involving several molecules (such as a crosslinking protein adhered to F-actin) has shown competition between force-activated unbinding and conformational changes<sup>66</sup>. A current challenge is determining how these molecular processes are integrated to mediate complex phenomena such as mechanotransduction.

## Models of dynamic FAs

FAs offer a convenient system for understanding the relationship between dynamics and mechanotransduction that may be more generally applicable. From a mechanical perspective, FAs are dynamic, deformable links between an elastic ECM and the force-generating actin cytoskeleton<sup>13,41</sup>. Myosin-generated forces are transmitted through the actin cytoskeleton to FA proteins. The applied forces affect the dissociation rates of the FA proteins from integrin receptors, from each

## BOX 2

# Mechanical properties of the cytoskeleton

Materials such as rubber or polyacrylamide gels are elastic solids, meaning that they deform rapidly in response to applied stress but spring back to their original shape when the stress ends. By contrast, liquids such as water or honey flow in response to applied stress such that deformation increases irreversibly and linearly with time. Many materials, including cells<sup>77</sup> and *in vitro* mixtures of cytoskeletal components<sup>20</sup>, show behaviours between these two limits and are referred to as viscoelastic. The behaviour of many such materials more closely resembles elastic materials on short timescales and viscous liquids on long timescales. A common example is silly putty, which flows like liquid when slowly squeezed but bounces like an elastic ball when thrown against the floor. On a molecular level, viscoelasticity is due to the presence of stress relaxation, usually through bond dissociation. At times shorter than the dissociation time, stress cannot be relaxed and the materials act like elastic solids, whereas on longer timescales, the bonds dissociate and the materials flow. The details of viscoelasticity in cytoskeletal networks are still controversial, but it is likely that the dissociation of, or conformational changes in, cytoskeletal crosslinking proteins is involved<sup>77</sup>.

other and from F-actin. Forces are then transmitted through bound molecules to deform the ECM (Fig. 2a).

Several models that address how cells respond to the mechanical properties of the ECM propose that rigidity determines how quickly forces act on the integrin–actin linkages<sup>13,67–69</sup>. The models can be classified on the basis of whether rapidly applied forces increase or decrease adhesion turnover by promoting adhesion breakage or strengthening, respectively. Although the detailed predictions are distinct, in all cases FA kinetics are determined by the balance of protein association and dissociation. On soft substrates or in response to slowly applied forces, the on rates and off rates of molecules into and out of the adhesions, and the lifetimes of the whole adhesions, are moderate<sup>67,68</sup> (Fig. 2b). In models without FA strengthening, exposing cells to large, rapidly applied forces or plating cells on rigid substrata increases the dissociation of linker molecules, such as vinculin or talin, from the integrin or the actin (slip-bond behaviour). However, with large numbers of unoccupied sites, there is also rapid rebinding. This model leads to FAs with faster exchange of linker molecules and shorter whole adhesion lifetimes (Fig. 2c). In models with FA strengthening, applied force results in decreased protein dissociation (catch-bond behaviour) and/or a conformational change that induces protein recruitment<sup>67,69</sup>. Either way, force leads to large, reinforced FAs with slower exchange rates for linker molecules and longer lifetimes (Fig. 2d).

Notably, FA dynamics consistent with both classes of model have been observed<sup>153,68,70</sup>. In some cases, the difference is cell-type dependent. There is also evidence for spatial specificity within single cells, such that FA strengthening is restricted to the front of migrating cells<sup>71</sup>. This result makes intuitive sense, because if adhesions always strengthened under force, cells could not migrate. A polarized mechanism that strengthens adhesions at the front while allowing those at the rear to break under tension will produce forward movement when the cell contracts.

The polarized signalling pathways that determine whether adhesions strengthen or weaken under force are unknown. A recent study<sup>72</sup> using a biosensor that reports the tension across the FA protein vinculin has helped to shed some light on the mechanism. It showed that vinculin is under high tension in FAs that assemble, whereas it is under low tension in FAs that disassemble under cellular contractile force in migrating

cells. Furthermore, vinculin is required for adhesion strengthening under force. These results indicate that the pathways that determine adhesion strengthening versus weakening under force regulate whether the force is transmitted across vinculin or other linkages.

## A dynamic model of mechanotransduction

The examples listed above suggest that a dynamic treatment of mechanotransduction is necessary. A key concept is that applied forces can regulate the rates of biochemically detectable processes, such as protein unbinding and protein conformational changes. Although switch-like models emphasize the serial nature of the steps of mechanotransduction, a dynamic model shows a more integrated picture in which mechanotransmission, mechanotransduction and mechanoresponse are intimately related and can affect each other.

## Dynamic mechanotransmission

Broken linkages cannot transmit forces. Thus, the stability of the load-bearing subcellular structures dictates paths of force transmission and their duration. On short timescales (subsecond to tens of seconds), mechanotransmission is governed by the physics of force-activated bond dissociation<sup>63,64</sup>. Whereas some cellular structures may simply be strong enough to bear the relevant forces, others will not. For example, bonds between actin and its crosslinking protein  $\alpha$ -actinin are slip bonds, and other calponin-homology-domain actin-binding proteins are likely to behave similarly<sup>73</sup>. Other crucial linkages in mechanotransduction, such as fibronectin–integrin ( $\alpha_5\beta_1$  integrin)<sup>74,75</sup> and actin–myosin<sup>76</sup> bonds, show catch-bond behaviour. These considerations suggest that only certain dynamic, subcellular structures may be stabilized in response to applied force to allow force transmission to mechanosensitive areas or molecules.

On larger length scales, the dynamic nature of cytoskeletal protein–protein bonds directly leads to viscoelasticity<sup>77</sup> (Box 2). Applied forces can result in either reinforcement<sup>23</sup> or fluidization<sup>78</sup> of the cytoskeleton. The exact mechanisms are still debated, but reinforcement is associated with the maintenance of physical linkages, stiffening of the actin network and increased cell contractility<sup>23,79</sup>. Fluidization involves disruption of the cytoskeleton, from either breakage of mechanical linkages<sup>80,81</sup> or force-induced, biochemically controlled disassembly<sup>82</sup>. In terms of mechanotransmission, these properties are extremely important as forces will be propagated along reinforced, elastic filaments, but quickly dissipate in a fluidized, viscous environment. Furthermore, viscoelastic effects can allow certain frequencies of mechanical stimulus to be selectively transmitted over greater distance in cells<sup>83</sup>. The efficiency of transmission for different frequencies is determined by the rates of bond dissociation that cause cellular viscoelasticity. These effects have been observed in force-induced movements of FAs<sup>84</sup> and mitochondria<sup>85</sup>. Mechanical stimuli with frequencies that are transmitted efficiently are likely to promote greater mechanoresponses.

## Dynamic mechanosensing

The basis of mechanosensing is thought to be force-sensitive changes in the rates of conversion between different protein conformations. These transitions depend on the strength and duration of force application (Box 1). For instance, when forces are applied to talin, 50 pN induces conformational changes within 25 ms, whereas at 20 pN, the same changes require 200 ms<sup>34</sup>. For successful mechanosensing, forces must be transmitted for sufficient time to induce conformational changes and subsequent biochemical detection. But force can also accelerate slip-bond breakage, which will terminate the force transmission. Thus, there is competition between conformational change and bond breakage. Transmission pathways with catch bonds will therefore be more sensitive.

The rate at which forces are applied influences force transmission and subsequent signalling. The rate of force application through F-actin to the actin-crosslinking proteins  $\alpha$ -actinin or filamin has been shown to determine the relative frequency of dissociation of the actin–linker bonds versus conformational changes<sup>66</sup>. Conformational changes



are more likely to occur at higher rates of force application. Other experiments have shown that fibronectin-coated beads are less likely to dissociate from integrins when forces are applied quickly<sup>86</sup>. The crucial effect of the rate of force application underscores the importance of these dynamic aspects of mechanotransduction.

### Dynamic mechanoresponse

The downstream mechanoresponse pathways are not innately force sensitive but often regulate cytoskeletal and adhesion structures that therefore feed back to influence mechanotransduction. The cytoskeletal protein zyxin, for instance, specifically localizes to areas of strain-induced stress-fibre thinning, and recruits  $\alpha$ -actinin and vasodilator-stimulated phosphoprotein, which promote actin polymerization and stress-fibre repair<sup>87,88</sup>. The myocardin-related transcription factor (also known as MAL) pathway is activated by actin polymerization in response to force or other stimuli, and regulates the expression of numerous cytoskeletal genes, including those encoding vinculin, filamin and actin<sup>42,43</sup>. Cyclic strain also enhances the expression of ECM proteins and induces the assembly of ECM structures<sup>89</sup>. Thus, on timescales of the order of minutes to days, cells use signalling or transcriptional programs to alter or maintain force-transmission pathways.

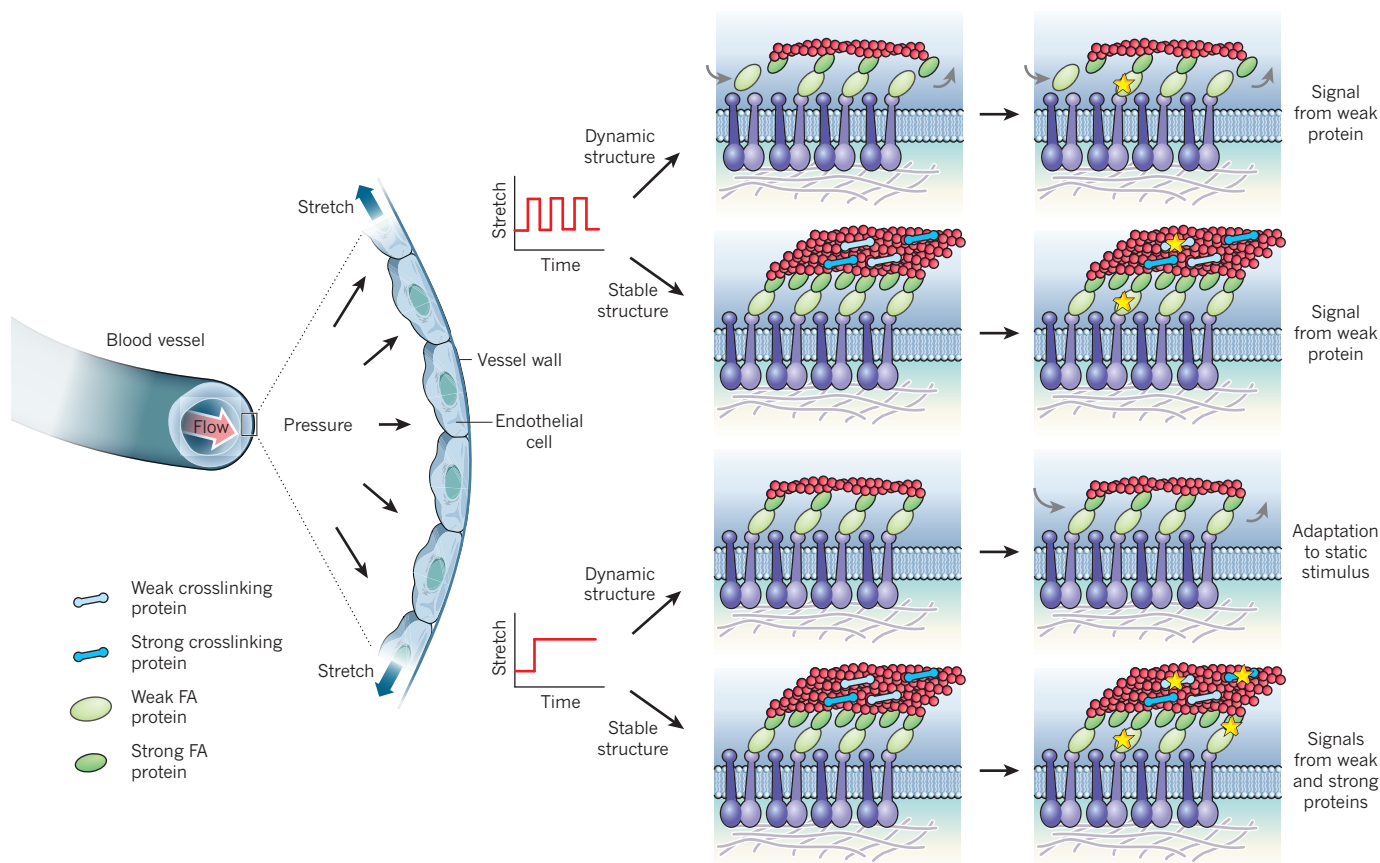
Cell alignment in response to applied force is a form of adaptation that involves local regulation of dynamic cytoskeletal elements, largely through the regulation of Rho family GTPases<sup>90</sup>. In two-dimensional cultures, uniaxial static stretch ('stretch and hold') induces actin stress fibre and FA

alignment parallel to the applied force, consistent with the general notion of adhesion strengthening<sup>90</sup>. By contrast, cyclic stretch induces alignment perpendicular to the applied force<sup>91</sup> in a frequency-dependent manner<sup>49</sup>.

A mathematical model recently proposed that forces applied faster than the characteristic rates of remodelling in load-bearing subcellular structures induce cell alignment perpendicular to the direction of strain to minimize stretching of these elements<sup>92</sup>. By contrast, when forces are applied slower than the remodelling rate, cells can internally remodel and align parallel to the applied stress. This concept may also explain a fascinating effect in which inhibiting Rho kinase or the Rho effector protein mammalian diaphanous has been found to shift the direction that cells align under cyclic stretch from perpendicular to parallel<sup>93</sup>. Inhibiting Rho signalling is also known to deplete cells of stable FAs and stress fibres, resulting in more dynamic subcellular structures<sup>41</sup>. The switch in direction of alignment may be explained if the higher cytoskeletal-remodelling rate now exceeds the characteristic rate of the cyclic stretch, which, according to the mathematical model, would yield alignment in the direction of strain. This model provides insight into how the internal dynamics of the cytoskeleton can determine responses to dynamic mechanical stimuli.

### Adhesion strengthening and rigidity sensing

The principles enumerated above can provide at least a first explanation for how cells sense the mechanical properties of their substrata (Fig. 2). A key point is that forces on ECM–integrin–cytoskeletal linkages



**Figure 3 | Dynamic aspects of mechanotransduction.** The endothelial cells that line blood-vessel walls are subject to both cyclic stretch (upper) and static stretch (lower). Even when the signal strengths are matched, dynamically distinct mechanical stimuli can activate common or unique signalling pathways based on the strength of the proteins — specifically the resistance to conformational changes — and the dynamic nature of the structures bearing loads. In dynamic structures, such as nascent adhesions, and stable structures, such as mature adhesions and stress fibres, cyclic stretch does not apply forces for sufficient time to induce conformational changes in strong proteins, but

weak proteins will signal (yellow star). In response to static stretch, dynamic structures can readily adapt, and there is no long-term signalling. In stable structures, the long force application causes conformational changes in both weak and strong proteins. Thus, signalling pathways that are preferentially activated by cyclic stretch are probably induced by weak proteins that localize exclusively to dynamic structures. Pathways selectively activated by static stretch are likely to contain mechanosensors that are strong proteins in stable structures. Pathways activated by both types of signal probably involve weak proteins that localize to both dynamic and stable structures.

build up more rapidly in cells on rigid surfaces than on compliant surfaces<sup>67,68</sup>. These more rapidly applied forces are better at triggering conformational changes in cytoskeletal proteins than at causing bond dissociation<sup>66</sup>. Furthermore, if these linkages contain catch bonds — whose conversion to a high-affinity state is predicted to be enhanced at high loading rates<sup>94</sup> — they will be further stabilized. As a result, mechanotransmission will be more efficient and longer lived. As a consequence of domain unfolding under force, additional proteins are recruited to support crucial linkages. Adhesion strengthening will then occur through the mechanisms we describe above.

### Sensing dynamic applied forces

One set of physiologically important mechanotransduction events involves the stretching of artery walls by blood pressure. Hypertension increases static stretch, whereas cyclic pumping during the cardiac cycle causes time-varying stretch<sup>4</sup>, both of which lead to the transmission of forces to vascular cells and activation of many signalling pathways<sup>95</sup>. However, cyclic stretch and static stretch of the same amplitude (and probably similar force-application rates) induce both common and unique mechanoresponses<sup>46</sup>. For example, a 10% cyclic stretch of endothelial cells increases the expression of vascular endothelial growth factor receptor-2 (VEGFR-2) and the angiotensin receptor TIE-2, but not VEGFR-1 expression. By contrast, a 10% static stretch increases the expression of VEGFR-2 and VEGFR-1, but not TIE-2 (ref. 46). We propose that these various responses can be explained by the dynamic processes intrinsic to the FAs and the cytoskeleton that sense the applied force (Fig. 3). Responses selectively activated by static stretch are probably mediated by protein conformational changes in strong proteins in relatively stable structures, requiring long applications of force to unfold. Responses selectively activated by cyclic stretch probably involve weaker proteins in dynamic structures that adapt to the statically applied forces but are constantly stimulated by dynamic signals. Responses that are activated by both are probably mediated by weak proteins in relatively stable structures.

### Future perspectives

The ability of mechanical perturbations to influence cellular signalling in a frequency-dependent manner can be conceptualized as mechanotransducers that function as bandpass filters, which selectively transmit specific frequencies. The ability of the cytoskeleton to transmit certain frequencies of mechanical stimuli to subcellular structures selectively provides one such mechanism<sup>83–85</sup>. Mechanosensitive elements and mechanoresponse pathways are also rate sensitive and frequency sensitive, owing to their own intrinsic timescales. In this regard, the timescale of the applied force must match the crucial timescale of a given signalling process to affect it. Stimuli that change too quickly are simply averaged, whereas stimuli that vary too slowly are not detected at all. Knowledge of the dynamics of cellular mechanotransducers could therefore enhance our understanding of frequency-dependent cell and tissue responses. As cells contain several mechanically sensitive biochemical signalling pathways with wide variations in important timescales, they may act as multiband pass filters, which pass several ranges of frequency. These systems would allow cells to distinguish multiple stimuli based on their frequencies or timescales.

Our understanding of mechanical signalling is still slim compared with our understanding of signalling by hormones and growth factors. But the more we learn, the more it seems that mechanical forces can have subtle and precise roles in governing morphogenesis, physiology and disease. We propose that just as conventional signals from soluble regulators act together in regulatory networks in which complex temporal and spatial characteristics determine outputs, so mechanical stresses may also convey large amounts of information through precise time-dependent and force-dependent modulation. For periodic stimuli, this will take the form of frequency and amplitude features that determine cellular outputs. Elucidating the dynamics of cellular mechanotransduction systems holds the key to understanding these mechanisms. ■

- Orr, A. W., Helmke, B. P., Blackman, B. R. & Schwartz, M. A. Mechanisms of mechanotransduction. *Dev. Cell* **10**, 11–20 (2006).
- Chen, C. S. Mechanotransduction — a field pulling together? *J. Cell Sci.* **121**, 3285–3292 (2008).
- Geiger, B., Spatz, J. P. & Bershadsky, A. D. Environmental sensing through focal adhesions. *Nature Rev. Mol. Cell Biol.* **10**, 21–33 (2009).
- Hahn, C. & Schwartz, M. A. Mechanotransduction in vascular physiology and atherogenesis. *Nature Rev. Mol. Cell Biol.* **10**, 53–62 (2009).
- Lucitti, J. L. *et al.* Vascular remodeling of the mouse yolk sac requires hemodynamic force. *Development* **134**, 3317–3326 (2007).
- Yashiro, K., Shiratori, H. & Hamada, H. Haemodynamics determined by a genetic programme govern asymmetric development of the aortic arch. *Nature* **450**, 285–288 (2007).
- Discher, D. E., Janmey, P. & Wang, Y. L. Tissue cells feel and respond to the stiffness of their substrate. *Science* **310**, 1139–1143 (2005).
- Levental, K. R. *et al.* Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**, 891–906 (2009).
- Liu, F. *et al.* Feedback amplification of fibrosis through matrix stiffening and COX-2 suppression. *J. Cell Biol.* **190**, 693–706 (2010).
- Engler, A. J., Sen, S., Sweeney, H. L. & Discher, D. E. Matrix elasticity directs stem cell lineage specification. *Cell* **126**, 677–689 (2006).
- Gilbert, P. M. *et al.* Substrate elasticity regulates skeletal muscle stem cell self-renewal in culture. *Science* **329**, 1078–1081 (2010).
- Dado, D. & Levenberg, S. Cell-scaffold mechanical interplay within engineered tissue. *Semin. Cell Dev. Biol.* **20**, 656–664 (2009).
- De, R., Zemel, A. & Safran, S. A. Theoretical concepts and models of cellular mechanosensing. *Methods Cell Biol.* **98**, 143–175 (2010).
- Saratzis, A. *et al.* Abdominal aortic aneurysm: a review of the genetic basis. *Angiology* **62**, 18–32 (2011).
- Hershberger, R. E., Morales, A. & Siegfried, J. D. Clinical and genetic issues in dilated cardiomyopathy: a review for genetics professionals. *Genet. Med.* **12**, 655–667 (2010).
- Chandrasekharan, K. & Martin, P. T. Genetic defects in muscular dystrophy. *Methods Enzymol.* **479**, 291–322 (2010).
- Laurent, S., Boutouyrie, P. & Lacolley, P. Structural and genetic bases of arterial stiffness. *Hypertension* **45**, 1050–1055 (2005).
- Yu, H., Mouw, J. K. & Weaver, V. M. Forcing form and function: biomechanical regulation of tumor evolution. *Trends Cell Biol.* **21**, 47–56 (2011).
- Schwartz, M. A. & DeSimone, D. W. Cell adhesion receptors in mechanotransduction. *Curr. Opin. Cell Biol.* **20**, 551–556 (2008).
- Gardel, M. L., Kasza, K. E., Brangwynne, C. P., Liu, J. & Weitz, D. A. Chapter 19 Mechanical response of cytoskeletal networks. *Methods Cell Biol.* **89**, 487–519 (2008).
- Helmke, B. P., Rosen, A. B. & Davies, P. F. Mapping mechanical strain of an endogenous cytoskeletal network in living endothelial cells. *Biophys. J.* **84**, 2691–2699 (2003).
- Tzima, E. *et al.* A mechanosensory complex that mediates the endothelial cell response to fluid shear stress. *Nature* **437**, 426–431 (2005).
- This paper identifies a crucial complex, consisting of PECAM1, VE-cadherin and VEGFR-2, in the pathway leading to integrin activation by shear flow.**
- Matthews, B. D., Overby, D. R., Mannix, R. & Ingber, D. E. Cellular adaptation to mechanical stress: role of integrins, Rho, cytoskeletal tension and mechanosensitive ion channels. *J. Cell Sci.* **119**, 508–518 (2006).
- Na, S. *et al.* Rapid signal transduction in living cells is a unique feature of mechanotransduction. *Proc. Natl Acad. Sci. USA* **105**, 6626–6631 (2008).
- Wang, N. & Stamenovic, D. Contribution of intermediate filaments to cell stiffness, stiffening, and growth. *Am. J. Physiol. Cell Physiol.* **279**, C188–C194 (2000).
- Hayakawa, K., Tatsumi, H. & Sokabe, M. Actin stress fibers transmit and focus force to activate mechanosensitive channels. *J. Cell Sci.* **121**, 496–503 (2008).
- Poh, Y. C. *et al.* Rapid activation of Rac GTPase in living cells by force is independent of Src. *PLoS ONE* **4**, e7886 (2009).
- Sukharev, S., Betanzos, M., Chiang, C. S. & Guy, H. R. The gating mechanism of the large mechanosensitive channel MscL. *Nature* **409**, 720–724 (2001).
- Árnadóttir, J. & Chalfie, M. Eukaryotic mechanosensitive channels. *Annu. Rev. Biophys.* **39**, 111–137 (2010).
- Zhong, C. *et al.* Rho-mediated contractility exposes a cryptic site in fibronectin and induces fibronectin matrix assembly. *J. Cell Biol.* **141**, 539–551 (1998).
- Oberhauser, A. F., Badilla-Fernandez, C., Carrion-Vazquez, M. & Fernandez, J. M. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J. Mol. Biol.* **319**, 433–447 (2002).
- Smith, M. L. *et al.* Force-induced unfolding of fibronectin in the extracellular matrix of living cells. *PLoS Biol.* **5**, e268 (2007).
- Ziegler, W. H., Gingras, A. R., Critchley, D. R. & Emsley, J. Integrin connections to the cytoskeleton through talin and vinculin. *Biochem. Soc. Trans.* **36**, 235–239 (2008).
- del Rio, A. *et al.* Stretching single talin rod molecules activates vinculin binding. *Science* **323**, 638–641 (2009).
- This study demonstrates force-induced binding of vinculin to cryptic sites in talin at the single molecule level.**
- Zhang, X. *et al.* Talin depletion reveals independence of initial cell spreading from integrin activation and traction. *Nature Cell Biol.* **10**, 1062–1068 (2008).
- Defilippi, P., Di Stefano, P. & Cabodi, S. p130Cas: a versatile scaffold in signaling networks. *Trends Cell Biol.* **16**, 257–263 (2006).
- Tamada, M., Sheetz, M. P. & Sawada, Y. Activation of a signaling cascade by cytoskeleton stretch. *Dev. Cell* **7**, 709–718 (2004).



38. Sawada, Y. *et al.* Rap1 is involved in cell stretching modulation of p38 but not ERK or JNK MAP kinase. *J. Cell Sci.* **114**, 1221–1227 (2001).
39. Sawada, Y. & Sheetz, M. P. Force transduction by Triton cytoskeletons. *J. Cell Biol.* **156**, 609–615 (2002).
40. Sawada, Y. *et al.* Force sensing by mechanical extension of the Src family kinase substrate p130Cas. *Cell* **127**, 1015–1026 (2006).  
**This study shows that stretching of p130<sup>Cas</sup> leads to the exposure of tyrosine residues, which can be phosphorylated to affect signalling pathways.**
41. Parsons, J. T., Horwitz, A. R. & Schwartz, M. A. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Rev. Mol. Cell Biol.* **11**, 633–643 (2010).
42. Asparuhova, M. B., Gelman, L. & Chiquet, M. Role of the actin cytoskeleton in tuning cellular responses to external mechanical stress. *Scand. J. Med. Sci. Sports* **19**, 490–499 (2009).
43. Olson, E. N. & Nordheim, A. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nature Rev. Mol. Cell Biol.* **11**, 353–365 (2010).
44. Mallion, J. M., Baguet, J. P., Siche, J. P., Tremel, F. & De Gaudemaris, R. Left ventricular hypertrophy and arterial hypertrophy. *Adv. Exp. Med. Biol.* **432**, 123–133 (1997).
45. Vogel, V. Mechanotransduction involving multimodular proteins: converting force into biochemical signals. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 459–488 (2006).
46. Zheng, W., Christensen, L. P. & Tomanek, R. J. Differential effects of cyclic and static stretch on coronary microvascular endothelial cell receptors and vasculogenic/angiogenic responses. *Am. J. Physiol. Heart Circ. Physiol.* **295**, H794–H800 (2008).  
**This study provides evidence that statically and dynamically applied stretches lead to the activation of distinct pathways in stretched endothelial cells.**
47. Lehoux, S., Esposito, B., Merval, R. & Tedgui, A. Differential regulation of vascular focal adhesion kinase by steady stretch and pulsatility. *Circulation* **111**, 643–649 (2005).
48. Hsu, H. J., Lee, C. F. & Kaunas, R. A dynamic stochastic model of frequency-dependent stress fiber alignment induced by cyclic stretch. *PLoS ONE* **4**, e4853 (2009).
49. Liu, B. *et al.* Role of cyclic strain frequency in regulating the alignment of vascular smooth muscle cells *in vitro*. *Biophys. J.* **94**, 1497–1507 (2008).
50. Gelfand, B. D., Epstein, F. H. & Blackman, B. R. Spatial and spectral heterogeneity of time-varying shear stress profiles in the carotid bifurcation by phase-contrast MRI. *J. Magn. Reson. Imaging* **24**, 1386–1392 (2006).
51. Dancu, M. B. & Tarbell, J. M. Large negative stress phase angle (SPA) attenuates nitric oxide production in bovine aortic endothelial cells. *J. Biomech. Eng.* **128**, 329–334 (2006).
52. Wehrle-Haller, B. Analysis of integrin dynamics by fluorescence recovery after photobleaching. *Methods Mol. Biol.* **370**, 173–201 (2007).
53. Hu, K., Ji, L., Applegate, K. T., Danuser, G. & Waterman-Storer, C. M. Differential transmission of actin motion within focal adhesions. *Science* **315**, 111–115 (2007).
54. Brown, C. M. *et al.* Probing the integrin–actin linkage using high-resolution protein velocity mapping. *J. Cell Sci.* **119**, 5204–5214 (2006).
55. Maruthamuthu, V., Aratyn-Schaus, Y. & Gardel, M. L. Conserved F-actin dynamics and force transmission at cell adhesions. *Curr. Opin. Cell Biol.* **22**, 583–588 (2010).
56. Liu, Z. *et al.* Mechanical tugging force regulates the size of cell–cell junctions. *Proc. Natl Acad. Sci. USA* **107**, 9944–9949 (2010).
57. Maruthamuthu, V., Sabass, B., Schwarz, U. S. & Gardel, M. L. Cell–ECM traction force modulates endogenous tension at cell–cell contacts. *Proc. Natl Acad. Sci. USA* **108**, 4708–4713 (2011).
58. Mège, R. M., Gavard, J. & Lambert, M. Regulation of cell–cell junctions by the cytoskeleton. *Curr. Opin. Cell Biol.* **18**, 541–548 (2006).
59. Ladoux, B. *et al.* Strength dependence of cadherin-mediated adhesions. *Biophys. J.* **98**, 534–542 (2010).
60. Kametani, Y. & Takeichi, M. Basal-to-apical cadherin flow at cell junctions. *Nature Cell Biol.* **9**, 92–98 (2007).
61. Riveline, D. *et al.* Focal contacts as mechanosensors: externally applied local mechanical force induces growth of focal contacts by an mDia1-dependent and ROCK-independent mechanism. *J. Cell Biol.* **153**, 1175–1186 (2001).
62. le Duc, Q. *et al.* Vinculin potentiates E-cadherin mechanosensing and is recruited to actin-anchored sites within adherens junctions in a myosin II-dependent manner. *J. Cell Biol.* **189**, 1107–1115 (2010).
63. Evans, E. A. & Calderwood, D. A. Forces and bond dynamics in cell adhesion. *Science* **316**, 1148–1153 (2007).
64. Thomas, W. E., Vogel, V. & Sokurenko, E. Biophysics of catch bonds. *Annu. Rev. Biophys.* **37**, 399–416 (2008).
65. Bustamante, C., Chemla, Y. R., Forde, N. R. & Izhaky, D. Mechanical processes in biochemistry. *Annu. Rev. Biochem.* **73**, 705–748 (2004).
66. Ferrer, J. M. *et al.* Measuring molecular rupture forces between single actin filaments and actin-binding proteins. *Proc. Natl Acad. Sci. USA* **105**, 9221–9226 (2008).
67. Bruinsma, R. Theory of force regulation by nascent adhesion sites. *Biophys. J.* **89**, 87–94 (2005).
68. Chan, C. E. & Odde, D. J. Traction dynamics of filopodia on compliant substrates. *Science* **322**, 1687–1691 (2008).  
**This study proposes and validates a model describing rigidity sensitive FA dynamics in terms of force-activated protein dissociation.**
69. Li, Y., Bhimalapuram, P. & Dinner, A. R. Model for how retrograde actin flow regulates adhesion traction stresses. *J. Phys. Condens. Matter* **22**, 194113 (2010).
70. Gardel, M. L. *et al.* Traction stress in focal adhesions correlates biphasically with actin retrograde flow speed. *J. Cell Biol.* **183**, 999–1005 (2008).
71. Schmidt, C. E., Horwitz, A. F., Lauffenburger, D. A. & Sheetz, M. P. Integrin–cytoskeletal interactions in migrating fibroblasts are dynamic, asymmetric, and regulated. *J. Cell Biol.* **123**, 977–991 (1993).
72. Grashoff, C. *et al.* Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature* **466**, 263–266 (2010).  
**This paper reports a biosensor that measures forces across specific proteins in dynamic FAs and shows that molecular tension across vinculin correlates with FA strengthening.**
73. Miyata, H., Yasuda, R. & Kinosita, K. Jr. Strength and lifetime of the bond between actin and skeletal muscle  $\alpha$ -actinin studied with an optical trapping technique. *Biochim. Biophys. Acta* **1290**, 83–88 (1996).
74. Kong, F., Garcia, A. J., Mould, A. P., Humphries, M. J. & Zhu, C. Demonstration of catch bonds between an integrin and its ligand. *J. Cell Biol.* **185**, 1275–1284 (2009).  
**This study shows that the linkage between  $\alpha_5\beta_1$  integrin and fibronectin acts like a catch bond at the single molecule level.**
75. Friedland, J. C., Lee, M. H. & Boettiger, D. Mechanically activated integrin switch controls  $\alpha_5\beta_1$  function. *Science* **323**, 642–644 (2009).  
**This study shows that force and increased extracellular rigidity switch  $\alpha_5\beta_1$  integrin between a relaxed and a tensioned state that is required for mechanically induced focal adhesion kinase signalling.**
76. Guo, B. & Guilford, W. H. Mechanics of actomyosin bonds in different nucleotide states are tuned to muscle contraction. *Proc. Natl Acad. Sci. USA* **103**, 9844–9849 (2006).
77. Hoffman, B. D. & Crocker, J. C. Cell mechanics: dissecting the physical responses of cells to force. *Annu. Rev. Biomed. Eng.* **11**, 259–288 (2009).
78. Trepat, X. *et al.* Universal physical responses to stretch in the living cell. *Nature* **447**, 592–595 (2007).
79. Gardel, M. L. *et al.* Prestressed F-actin networks cross-linked by hinged filamins replicate mechanical properties of cells. *Proc. Natl Acad. Sci. USA* **103**, 1762–1767 (2006).
80. Lee, H., Ferrer, J. M., Lang, M. J. & Kamm, R. D. Molecular origin of strain softening in cross-linked F-actin networks. *Phys. Rev. E* **82**, 011919 (2010).
81. Chaudhuri, O., Parekh, S. H. & Fletcher, D. A. Reversible stress softening of actin networks. *Nature* **445**, 295–298 (2007).
82. Chen, C. *et al.* Fluidization and solidification of the human bladder smooth muscle cell in response to transient stretch. *PLoS ONE* **5**, e12035 (2010).
83. Shafir, Y. & Forgacs, G. Mechanotransduction through the cytoskeleton. *Am. J. Physiol. Cell Physiol.* **282**, C479–C486 (2002).
84. Mack, P. J., Kaazempur-Mofrad, M. R., Karcher, H., Lee, R. T. & Kamm, R. D. Force-induced focal adhesion translocation: effects of force amplitude and frequency. *Am. J. Physiol. Cell Physiol.* **287**, C954–C962 (2004).
85. Hu, S. & Wang, N. Control of stress propagation in the cytoplasm by prestress and loading frequency. *Mol. Cell. Biomech.* **3**, 49–60 (2006).
86. Jiang, G., Huang, A. H., Cai, Y., Tanase, M. & Sheetz, M. P. Rigidity sensing at the leading edge through  $\alpha_5\beta_3$  integrins and RPTP $\alpha$ . *Biophys. J.* **90**, 1804–1809 (2006).
87. Smith, M. A. *et al.* A zyxin-mediated mechanism for actin stress fiber repair. *Dev. Cell* **19**, 365–376 (2010).
88. Wojtowicz, A. *et al.* Zyxin mediation of stretch-induced gene expression in human endothelial cells. *Circ. Res.* **107**, 898–902 (2010).
89. Chiquet, M., Gelman, L., Lutz, R. & Maier, S. From mechanotransduction to extracellular matrix gene expression in fibroblasts. *Biochim. Biophys. Acta* **1793**, 911–920 (2009).
90. Katsumi, A. *et al.* Effects of cell tension on the small GTPase Rac. *J. Cell Biol.* **158**, 153–164 (2002).
91. Kanda, K. & Matsuda, T. Behavior of arterial-wall cells cultured on periodically stretched substrates. *Cell Transplant.* **2**, 475–484 (1993).
92. De, R., Zemel, A. & Safran, S. A. Dynamics of cell orientation. *Nature Phys.* **3**, 655–659 (2007).  
**This theory-based study suggests how cytoskeletal dynamics affect the ability of cells to align to dynamically applied stretches.**
93. Kaunas, R., Nguyen, P., Usami, S. & Chien, S. Cooperative effects of Rho and mechanical stretch on stress fiber organization. *Proc. Natl Acad. Sci. USA* **102**, 15895–15900 (2005).
94. Prezhdov, O. V. & Pereverzev, Y. V. Theoretical aspects of the biological catch bond. *Acc. Chem. Res.* **42**, 693–703 (2009).
95. Haga, J. H., Li, Y. S. & Chien, S. Molecular basis of the effects of mechanical stretch on vascular smooth muscle cells. *J. Biomech.* **40**, 947–960 (2007).

**Acknowledgements** This work was supported by R01 grant HL075092 from the US Public Health Service to M.A.S. and by a fellowship from the American Heart Association to B.D.H.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to M.A.S. ([maschwartz@virginia.edu](mailto:maschwartz@virginia.edu)).

# Molecular chaperones in protein folding and proteostasis

F. Ulrich Hartl<sup>1</sup>, Andreas Bracher<sup>1</sup> & Manajit Hayer-Hartl<sup>1</sup>

**Most proteins must fold into defined three-dimensional structures to gain functional activity. But in the cellular environment, newly synthesized proteins are at great risk of aberrant folding and aggregation, potentially forming toxic species. To avoid these dangers, cells invest in a complex network of molecular chaperones, which use ingenious mechanisms to prevent aggregation and promote efficient folding. Because protein molecules are highly dynamic, constant chaperone surveillance is required to ensure protein homeostasis (proteostasis). Recent advances suggest that an age-related decline in proteostasis capacity allows the manifestation of various protein-aggregation diseases, including Alzheimer's disease and Parkinson's disease. Interventions in these and numerous other pathological states may spring from a detailed understanding of the pathways underlying proteome maintenance.**

Proteins are the most versatile and structurally complex biological macromolecules. They are involved in almost every biological process. Mammalian cells typically express in excess of 10,000 different protein species, which are synthesized on ribosomes as linear chains of up to several thousand amino acids. To function, these chains must generally fold into their 'native state', an ensemble of a few closely related three-dimensional structures<sup>1,2</sup>. How this is accomplished and how cells ensure the conformational integrity of their proteome in the face of acute and chronic challenges constitute one of the most fundamental and medically relevant problems in biology.

Central to this problem is that proteins must retain conformational flexibility to function, and thus are only marginally thermodynamically stable in their physiological environment. A substantial fraction of all proteins in eukaryotic cells (20–30% of the total in mammalian cells) even seem to be inherently devoid of any ordered three-dimensional structure and adopt folded conformations only after interaction with binding partners<sup>3</sup>. Aberrant behaviour of some of these metastable proteins, such as tau and  $\alpha$ -synuclein, can give rise to the formation of fibrillar aggregates that are associated with dementia and Parkinson's disease. Thus, protein quality control and the maintenance of proteome homeostasis (known as proteostasis) are crucial for cellular and organismal health. Proteostasis is achieved by an integrated network of several hundred proteins<sup>4</sup>, including, most prominently, molecular chaperones and their regulators, which assist in *de novo* folding or refolding, and the ubiquitin–proteasome system (UPS) and autophagy system, which mediate the timely removal of irreversibly misfolded and aggregated proteins. Deficiencies in proteostasis have been shown to facilitate the manifestation or progression of numerous diseases, such as neurodegeneration and dementia, type 2 diabetes, peripheral amyloidosis, lysosomal storage disease, cystic fibrosis, cancer and cardiovascular disease. A major risk factor for many of these ailments is advanced age. Indeed, studies in model organisms indicate that ageing is linked to a gradual decline in cellular proteostasis capacity<sup>5,6</sup>.

Here we discuss recent insights into the mechanisms of chaperone-assisted protein folding and proteome maintenance. We focus on how proteins use the chaperone machinery to navigate successfully the complex folding-energy landscape in the crowded cellular environment. Understanding these reactions will guide future efforts to define the proteostasis network as a target for pharmacological intervention in diseases of aberrant protein folding.

## Fundamental role of molecular chaperones

Many small proteins refold after their removal from denaturant *in vitro*, in the absence of other components or an energy source. This signifies that the amino-acid sequence, encoded in the DNA, contains all of the necessary information to specify the three-dimensional structure of a protein<sup>1</sup>. However, research over the past couple of decades has firmly established that in the cellular environment, many proteins require molecular chaperones to fold efficiently and on a biologically relevant timescale<sup>7</sup>. Why is this extra layer of complexity necessary?

Although small proteins may fold at very fast speeds<sup>8</sup> (within microseconds), in dilute buffer solutions, larger, multidomain proteins may take minutes to hours to fold<sup>9</sup>, and often even fail to reach their native states *in vitro*. The folding of such proteins becomes considerably more challenging *in vivo*, because the cellular environment is highly crowded, with total cytosolic protein reaching concentrations of 300–400 g l<sup>-1</sup>. The resultant excluded volume effects, although enhancing the functional interactions between macromolecules, also strongly increase the tendency of non-native and structurally flexible proteins to aggregate<sup>10</sup>. It seems likely, therefore, that the fundamental requirement for molecular chaperones arose very early during the evolution of densely crowded cells, owing to the need to minimize protein aggregation during folding and maintain proteins in soluble, yet conformationally dynamic states. Moreover, as mutations often disrupt the ability of a protein to adopt a stable fold<sup>11</sup>, it follows that the chaperone system provides a crucial buffer, allowing the evolution of new protein functions and phenotypic traits<sup>11,12</sup>.

## Some basics on protein folding and how it can go awry

Because the number of possible conformations a protein chain can adopt is very large, folding reactions are highly complex and heterogeneous, relying on the cooperation of many weak, non-covalent interactions. In the case of soluble proteins, hydrophobic forces are particularly important in driving chain collapse and the burial of non-polar amino-acid residues within the interior of the protein (see ref. 13 for a discussion of membrane protein folding). Considerable progress has been made in recent years in understanding these reactions through biophysical experiments and theoretical analyses<sup>1,2</sup>. In the current model, polypeptide chains are thought to explore funnel-shaped potential energy surfaces as they progress, along several

<sup>1</sup>Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

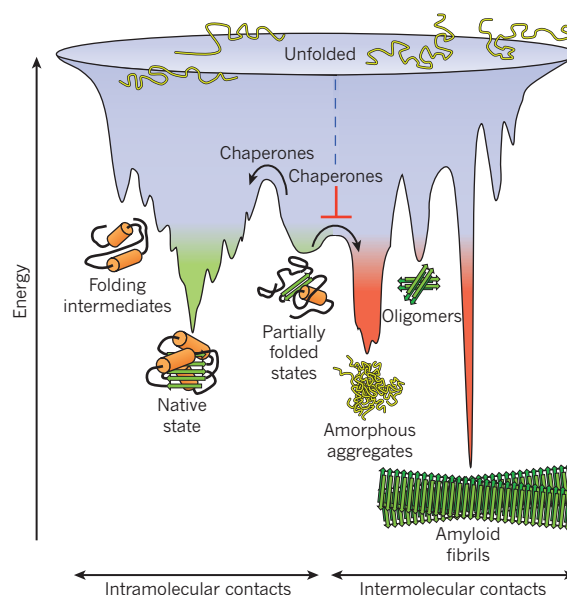


downhill routes, towards the native structure (Fig. 1). Chain collapse and the progressive increase in the number of native interactions rapidly restrict the conformational space that needs to be searched en route to the native state. However, the free-energy surface that must be navigated is often rugged, which means that the molecules must cross substantial kinetic barriers during folding. As a consequence, partially folded states may become transiently populated as kinetically trapped species. Such folding intermediates are the rule for proteins larger than 100 amino acids (~90% of all proteins in a cell), which have a strong tendency to undergo rapid hydrophobic collapse into compact globular conformations<sup>2</sup>. The collapse may lead either to disorganized globules lacking specific contacts and retaining large configurational entropy or to intermediates that may be stabilized by non-native interactions (misfolded states). In the former case, the search for crucial native contacts within the globule will limit folding speed, whereas in the latter, the breakage of non-native contacts may be rate-limiting<sup>1</sup> (Fig. 1). The propensity of proteins to populate globular intermediates with a high degree of flexibility may increase with larger, topologically more complex domain folds that are stabilized by many long-range interactions (such as  $\alpha/\beta$  domain architectures). Such proteins are often highly chaperone dependent<sup>14</sup>.

Partially folded or misfolded states are problematic because they tend to aggregate in a concentration-dependent manner (Fig. 1). This is due to the fact that these forms typically expose hydrophobic amino-acid residues and regions of unstructured polypeptide backbone to the solvent — features that become buried in the native state<sup>15</sup>. Like intramolecular folding, aggregation is largely driven by hydrophobic forces and primarily results in amorphous structures (Fig. 1). Alternatively, fibrillar aggregates called amyloid may form, defined by  $\beta$ -strands that run perpendicular to the long fibril axis (cross- $\beta$  structure). Although many proteins can adopt these highly ordered, thermodynamically stable structures under conditions *in vitro*<sup>16</sup>, the formation of these aggregates *in vivo* is strongly restricted by the chaperone machinery, suggesting that they may become more widespread under stress or when protein quality control fails. Importantly, the formation of fibrillar aggregates is often accompanied by the formation of soluble oligomeric states, which are thought to have key roles in diseases of aberrant folding<sup>16</sup> (Fig. 1). The toxicity of these less ordered and rather heterogeneous forms has been suggested to correlate with the exposure of sticky, hydrophobic surfaces and accessible peptide-backbone structure that is not yet integrated into a stable cross- $\beta$  core<sup>17</sup>. The soluble oligomers must undergo considerable rearrangement to form fibrils, the thermodynamic end state of the aggregation process, and may thus be comparable to the kinetically trapped intermediates in folding (Fig. 1). Notably, some common structural epitopes have been detected on the prefibrillar oligomers of different polypeptides<sup>18</sup>, but how these features are linked with toxicity is not yet understood. Such information is urgently needed to develop treatments for the numerous pathological states associated with protein aggregation.

### Major chaperone classes

We define a molecular chaperone as any protein that interacts with, stabilizes or helps another protein to acquire its functionally active conformation, without being present in its final structure<sup>7,19</sup>. Several different classes of structurally unrelated chaperones exist in cells, forming cooperative pathways and networks. Members of these protein families are often known as stress proteins or heat-shock proteins (HSPs), as they are upregulated under conditions of stress in which the concentrations of aggregation-prone folding intermediates increase. Chaperones are usually classified according to their molecular weight (HSP40, HSP60, HSP70, HSP90, HSP100 and the small HSPs). They are involved in a multitude of proteome-maintenance functions, including *de novo* folding, refolding of stress-denatured proteins, oligomeric assembly, protein trafficking and assistance in proteolytic degradation. The chaperones that participate broadly in *de novo* protein folding and refolding, such as the HSP70s, HSP90s and the chaperonins (HSP60s), are multicomponent molecular machines that promote folding through



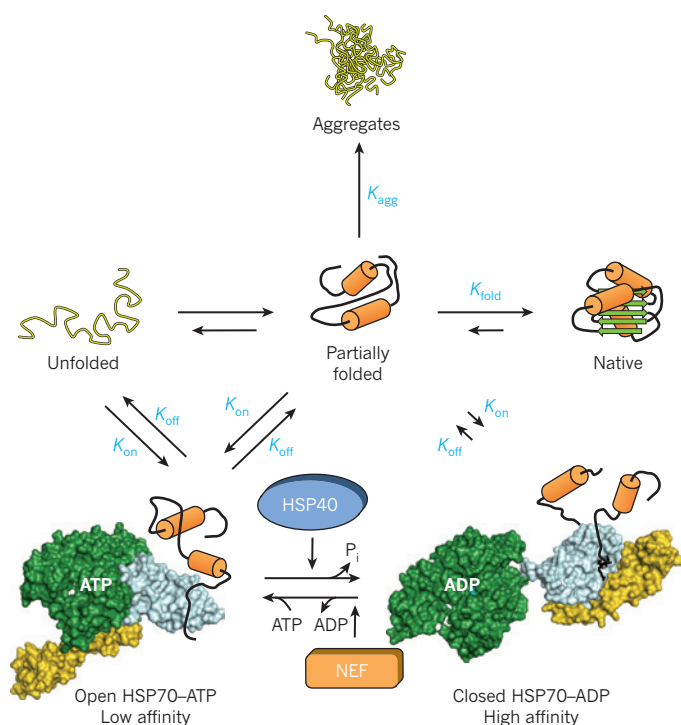
**Figure 1 | Competing reactions of protein folding and aggregation.** Scheme of the funnel-shaped free-energy surface that proteins explore as they move towards the native state (green) by forming intramolecular contacts (modified from refs 19 and 95). The ruggedness of the free-energy landscape results in the accumulation of kinetically trapped conformations that need to traverse free-energy barriers to reach a favourable downhill path. *In vivo*, these steps may be accelerated by chaperones<sup>39,41,42</sup>. When several molecules fold simultaneously in the same compartment, the free-energy surface of folding may overlap with that of intermolecular aggregation, resulting in the formation of amorphous aggregates, toxic oligomers or ordered amyloid fibrils (red). Fibrillar aggregation typically occurs by nucleation-dependent polymerization. It may initiate from intermediates populated during *de novo* folding or after destabilization of the native state (partially folded states) and is normally prevented by molecular chaperones.

ATP- and cofactor-regulated binding and release cycles. They typically recognize hydrophobic amino-acid side chains exposed by non-native proteins and may functionally cooperate with ATP-independent chaperones, such as the small HSPs, which function as ‘holdases’, buffering aggregation.

In the ATP-dependent mechanism of chaperone action, *de novo* folding and protein refolding is promoted through kinetic partitioning (Fig. 2). Chaperone binding (or rebinding) to hydrophobic regions of a non-native protein transiently blocks aggregation; ATP-triggered release allows folding to proceed. Importantly, although the HSP70s and the chaperonins both operate by this basic mechanism, they differ fundamentally in that the former (like all other ATP-dependent chaperones) release the substrate protein for folding into bulk solution, whereas the cylindrical chaperonins allow the folding of single protein molecules enclosed in a cage. The two systems act sequentially, whereby HSP70 interacts upstream with nascent and newly synthesized polypeptides and the chaperonins function downstream in the final folding of those proteins that fail to reach native state by cycling on HSP70 alone<sup>20,21</sup> (Figs 2 and 3). In the following sections, we will use the HSP70, chaperonin and HSP90 models to illustrate the basic mechanisms of the major cytosolic protein-folding machines. Client-specific chaperones that function downstream of folding in mediating the assembly of oligomeric complexes are not discussed (see, for example, refs 22 and 23).

### The HSP70 system

The constitutively expressed (HSC70, also known as HSPA8) and stress-inducible forms of HSP70 are central players in protein folding and proteostasis control. Increasing HSP70 levels has also proven effective in preventing toxic protein aggregation in disease models<sup>24</sup>.



**Figure 2 | The HSP70 chaperone cycle.** HSP70 is switched between high- and low-affinity states for unfolded and partially folded protein by ATP binding and hydrolysis. Unfolded and partially folded substrate (nascent chain or stress-denatured protein), exposing hydrophobic peptide segments, is delivered to ATP-bound HSP70 (open; low substrate affinity with high on-rates and off-rates) by one of several HSP40 cofactors. The hydrolysis of ATP, which is accelerated by HSP40, results in closing of the  $\alpha$ -helical lid of the peptide-binding domain (yellow) and tight binding of substrate by HSP70 (closed; high affinity with low on-rates and off-rates). Dissociation of ADP catalysed by one of several nucleotide-exchange factors (NEFs) is required for recycling. Opening of the  $\alpha$ -helical lid, induced by ATP binding, results in substrate release. Folding is promoted and aggregation is prevented when both the folding rate constant ( $K_{\text{fold}}$ ) is greater than the association constant ( $K_{\text{on}}$ ) for chaperone binding (or rebinding) of partially folded states, and  $K_{\text{on}}$  is greater than intermolecular association by the higher-order aggregation rate constant  $K_{\text{agg}}$  ( $K_{\text{fold}} > K_{\text{on}} > K_{\text{agg}}$ ) (kinetic partitioning). For proteins that populate misfolded states,  $K_{\text{on}}$  may be greater than  $K_{\text{fold}}$  ( $K_{\text{fold}} \leq K_{\text{on}} > K_{\text{agg}}$ ). These proteins are stabilized by HSP70 in a non-aggregated state, but require transfer into the chaperonin cage for folding<sup>14,20</sup>. After conformational stress,  $K_{\text{agg}}$  may become faster than  $K_{\text{on}}$ , and aggregation occurs ( $K_{\text{agg}} > K_{\text{on}} \geq K_{\text{fold}}$ ), unless chaperone expression is induced via the stress-response pathway. Structures in this figure relate to Protein Data Bank (PDB) accession codes 1DKG, 1DKZ, 2KHO and 2QXL. P<sub>i</sub>, inorganic phosphate.

The ATP-dependent reaction cycle of HSP70 is regulated by chaperones of the HSP40 (also known as DnaJ) family and nucleotide-exchange factors<sup>25,26</sup>. Some of these factors are also involved in linking chaperone functions with the UPS and autophagy for the removal of misfolded proteins<sup>27</sup>. Binding and release by HSP70 is achieved through the allosteric coupling of a conserved amino-terminal ATPase domain with a carboxy-terminal peptide-binding domain, the latter consisting of a  $\beta$ -sandwich subdomain and an  $\alpha$ -helical lid segment<sup>25</sup> (Fig. 2). The  $\beta$ -sandwich recognizes extended,  $\sim$ seven-residue segments enriched in hydrophobic amino acids, preferentially when they are framed by positively charged residues<sup>28</sup>. Such segments occur on average every 50–100 amino acids in proteins, and the exposure of these fragments correlates with the aggregation propensity of the protein<sup>29</sup>. The  $\alpha$ -helical lid and a conformational change in the  $\beta$ -sandwich domain regulate the affinity state for the peptide in an ATP-dependent manner<sup>25</sup>. In the ATP-bound state, the lid adopts an open conformation, resulting in high on

rates and off rates for the peptide. Hydrolysis of ATP to ADP is strongly accelerated by HSP40, leading to lid closure and stable peptide binding (low on rates and off rates for the peptide substrate) (Fig. 2). HSP40 also interacts directly with unfolded polypeptides and can recruit HSP70 to protein substrates<sup>20,26</sup>. After ATP hydrolysis, a nucleotide-exchange factor binds to the HSP70 ATPase domain and catalyses ADP–ATP exchange, resulting in lid opening and substrate release. Release allows fast-folding molecules to bury hydrophobic residues, whereas molecules that need longer than a few seconds for folding will rebinding to HSP70, thereby avoiding aggregation. HSP70 (re)binding may also result in conformational remodelling, perhaps removing kinetic barriers to the folding process<sup>30</sup>.

Proteins that are unable to partition to fast-folding trajectories after HSP70 cycling may be transferred into the specialized environment of the chaperonin cage for folding. Among these are several essential proteins, such as actins and tubulins<sup>31</sup>, which encounter high energetic barriers in folding and are completely unable to reach their native states spontaneously, even in dilute solution *in vitro*.

### The chaperonins

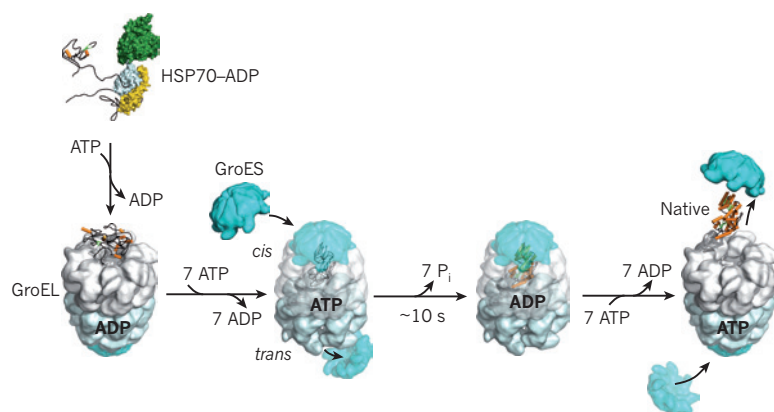
Chaperonins are large double-ring complexes of  $\sim$ 800–900 kDa that function by globally enclosing substrate proteins up to  $\sim$ 60 kDa for folding. Group I chaperonins (also known as HSP60s in eukaryotes and GroEL in bacteria) have seven-membered rings in bacteria, mitochondria and chloroplasts, and functionally cooperate with HSP10 proteins (GroES in bacteria), which form the lid of the folding cage. The group II chaperonins in archaea (thermosome) and the eukaryotic cytosol (TRiC, also known as CCT) usually have eight-membered rings. They are independent of HSP10 factors.

The GroEL–GroES chaperonin system of *Escherichia coli* has been studied most extensively<sup>19,32</sup> (Fig. 3). GroEL interacts with at least 250 different cytosolic proteins. Most of these are between 20 and 50 kDa in size and have complex  $\alpha/\beta$  or  $\alpha+\beta$  domain topologies, such as the TIM barrel fold<sup>14,33</sup>. These proteins are stabilized by many long-range interactions and are thought to populate flexible, kinetically trapped folding intermediates exposing hydrophobic surfaces<sup>34,35</sup>. The apical domains of GroEL present hydrophobic amino-acid residues for substrate binding in the ring centre. Subsequent folding depends on global substrate encapsulation by GroES (Fig. 3). GroEL binding is ATP regulated and is associated with a marked conformational change of GroEL that leads to the formation of a cage with a highly hydrophilic, net-negatively-charged inner wall<sup>19,32,36</sup>. Encapsulated protein is free to fold in this environment for  $\sim$ 10 seconds — the time needed for ATP hydrolysis in the GroES-bound ring (*cis* ring). Protein substrate leaves the cage after GroES dissociation, which is allosterically triggered by ATP binding in the opposite ring (*trans* ring). Not-yet folded substrate rapidly rebinding to GroEL for further folding attempts.

Enclosing unfolded protein, one molecule at a time, avoids disruption of folding by aggregation or (re)binding to upstream chaperones. In addition, an effect of steric confinement probably modulates the folding-energy landscape. Although the chaperonin functions as a passive-aggregation prevention device for some proteins<sup>32,37</sup>, encapsulation can also accelerate folding substantially<sup>37–39</sup>. This rate acceleration may be due to steric confinement, entropically destabilizing collapsed yet flexible folding intermediates, and promoting their conversion to more compact, native-like conformations. As shown recently, the effect of the folding cage may be comparable to the role of disulphide bonds in restricting conformational space in the folding of secretory proteins<sup>39</sup>. Furthermore, repeated unfolding events in successive binding and release cycles have been suggested to reverse misfolded, kinetically trapped states that are stabilized by non-native interactions<sup>40–42</sup>. Thus, the chaperonins may be able to remove both entropic and enthalpic barriers in rugged free-energy landscapes of folding (Fig. 1).

TRiC, the group II chaperonin in the eukaryotic cytosol, consists of eight paralogous subunits per ring<sup>31,43,44</sup>. All group II chaperonins deviate from GroEL in that their apical domains contain finger-like





**Figure 3 | Folding in the GroEL–GroES chaperonin cage.** Substrate binding to GroEL (after transfer from HSP70) may result in local unfolding<sup>42</sup>. ATP binding then triggers a conformational rearrangement of the GroEL apical domains. This is followed by the binding of GroES (forming the *cis* complex) and substrate encapsulation for folding. At the same time, ADP and GroES dissociate from the opposite (*trans*) GroEL ring, allowing the release of substrate that had been enclosed in the former *cis* complex<sup>32</sup> (omitted from the figure for simplicity). The new substrate remains encapsulated, free to fold, for the time needed to hydrolyse the seven ATP molecules in the newly formed *cis* complex (~10 s). Binding of ATP and GroES to the *trans* ring causes the opening of the *cis* complex. A symmetrical GroEL–(GroES)<sub>2</sub> complex may form transiently. Structural model is based on PDB accession 1AON.

protrusions, which act as an iris-like, built-in lid and replace the function of GroES. These segments open and close in an ATP-dependent protein-encapsulation cycle, similar in principle to that of GroEL–GroES<sup>44</sup>. However, the TRiC reaction cycle is much slower than that of GroEL, probably providing a substantially longer period of protein encapsulation and folding in the cage<sup>45</sup>. TRiC interacts with approximately 10% of newly synthesized cytosolic proteins, including actin and tubulins<sup>31,43</sup>. Interestingly, TRiC also functions in preventing the accumulation of toxic aggregates by the Huntington's disease protein<sup>46–48</sup>.

### The HSP90 system

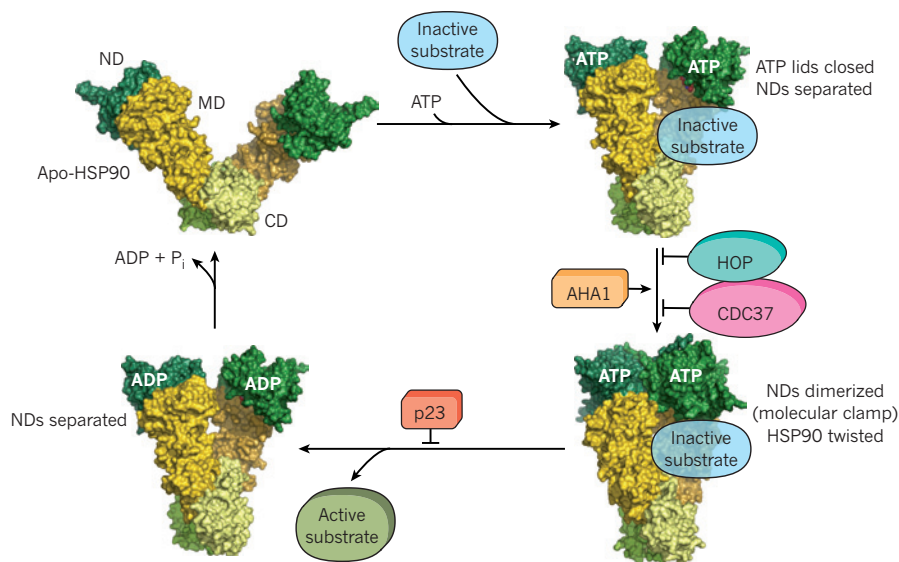
HSP90 forms a proteostasis hub that controls numerous important signalling pathways in eukaryotic cells<sup>49</sup>. These pleiotropic functions include, among others, cell-cycle progression, telomere maintenance, apoptosis, mitotic signal transduction, vesicle-mediated transport, innate immunity and targeted protein degradation. Indeed, the evolution and maintenance of these functional networks is thought to depend on the ability of HSP90 to buffer the effects of structurally destabilizing mutations in the underlying protein complexes, thereby allowing the acquisition of new traits<sup>12</sup>.

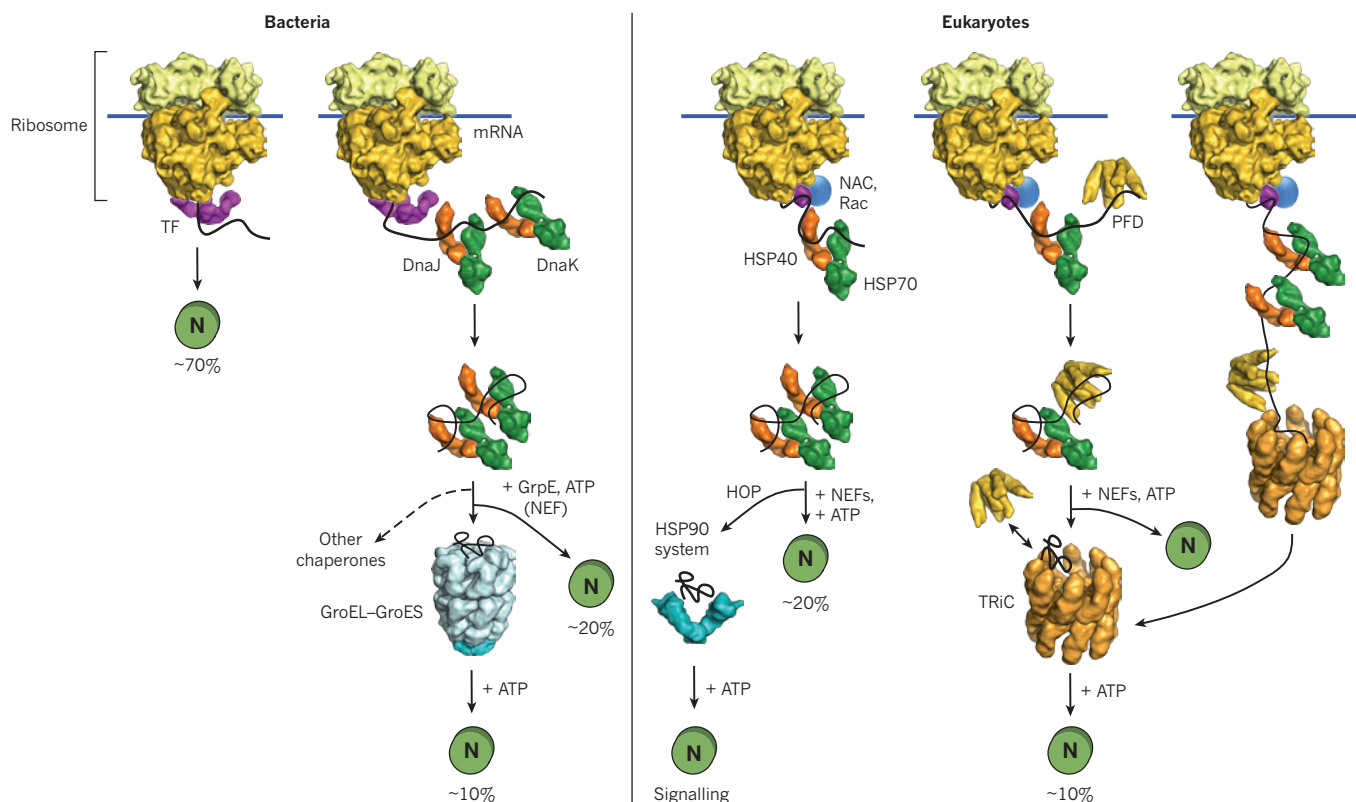
HSP90 functions downstream of HSP70 in the structural maturation and conformational regulation of numerous signal-transduction molecules, such as kinases and steroid receptors<sup>49,50</sup>. It cooperates in this process with several regulators and co-chaperones, many of which use tetratricopeptide repeat (TPR) domains to dock onto HSP90. For example, the TPR protein HOP provides a direct link between HSP70 and HSP90, allowing substrate transfer<sup>51</sup>. Although the mechanism by which HSP90 and its cofactors mediate conformational changes

in substrate proteins is not yet understood<sup>52</sup>, recent crystal structures of full-length HSP90s provided long-awaited information<sup>53,54</sup>. HSP90 functions as a dimer of subunits that are assembled by their C-terminal domains. An N-terminal domain binds and hydrolyses ATP and is joined to the C-terminal domain by a middle domain (Fig. 4). The middle domain participates in substrate binding and interacts with the co-chaperone AHA1. Similar to other chaperones, the HSP90 dimer undergoes an ATP-driven reaction cycle that is accompanied by considerable structural rearrangement<sup>25</sup> (Fig. 4). ATP binding leads to the dimerization of the N-terminal domains, forming the HSP90 'molecular clamp'. This results in a compaction of the HSP90 dimer, in which the individual monomers twist around each other. After hydrolysis, the ATPase domains dissociate, and the HSP90 monomers separate N-terminally. Various cofactors regulate this cycle: CDC37, which delivers certain kinase substrates to HSP90, inhibits the ATPase activity, and HOP inhibits N-terminal dimerization. AHA1 stimulates ATP hydrolysis, whereas p23 stabilizes the dimerized form of HSP90 before ATP hydrolysis. These factors are thought to adjust the kinetic properties of the cycle to achieve certain conformational transitions in HSP90-bound substrates, as well as their release from HSP90.

How HSP90 recruits different types of substrate protein with the help of various co-chaperones remains enigmatic. HSP90 appears to have several substrate-interaction regions, and the binding strength seems to be strongly influenced by the structural flexibility of the substrate<sup>52</sup>, in line with the proposed role of HSP90 as an evolutionary capacitor in protecting mutated protein variants from degradation<sup>12</sup>. Because several HSP90 substrates are kinases with well-documented roles in tumour development, the inhibition of HSP90 with drugs

**Figure 4 | ATPase cycle of the HSP90 chaperone system.** Clockwise from top left, ATP binding to the N-terminal ATPase domain (ND) of apo-HSP90 induces a conformational change and the closure of the ATP lid in the ND. After lid closure, the NDs dimerize, forming the closed HSP90 dimer (molecular clamp) with twisted subunits. This metastable conformation is committed for ATP hydrolysis. After hydrolysis, the NDs dissociate. The inactive substrate molecule interacts mostly with the middle domain (MD) and is conformationally activated as HSP90 proceeds through the ATPase cycle. The cofactors CDC37, HOP, AHA1 and p23 accelerate or slow certain steps of the cycle. Structures relate to PDB accessions 2IQO, 2O1U, 2CG9 and 2O1V. CD, C-terminal ATPase domain.





**Figure 5 | Organization of chaperone pathways in the cytosol.** In bacteria (left) and eukaryotes (right), chaperones that function in stabilizing nascent polypeptides on ribosomes and in initiating folding cooperate with machinery that acts downstream in completing folding<sup>7,19–21,31</sup>. The number of interacting substrates is indicated as a percentage of the total proteome. The first category of factors includes chaperones that bind in close proximity to the ribosomal polypeptide exit site, such as trigger factor (TF) in bacteria and specialized HSP70 complexes (ribosome-associated complex (Rac) in *Saccharomyces cerevisiae*, MPP11 and HSP70L1 in mammalian cells) and nascent-chain-associated complex (NAC) in eukaryotes). These chaperones bind hydrophobic

chain segments. Non-ribosome-bound members of the HSP70 family (DnaK in bacteria and HSC70 in eukaryotes) function as second-tier chaperones for longer nascent chains, mediating co- or post-translational folding. They also distribute subsets of proteins to downstream chaperones, such as the chaperonins (GroEL in bacteria and TRiC in eukaryotes)<sup>14,20,21</sup> and HSP90<sup>52</sup>. Substrate transfer from HSC70 to HSP90 is promoted by the coupling protein HOP. Dashed arrow indicates that pathway is not well established. Structures relate to PDB accessions 1DKG, 1DKZ, 2KHO, 1W26, 3IYF, 2AVY, 2AW4, 1FXK, 3LKX, 2IOQ, 2QWR and 1NLT. N, native protein; GrpE, protein GrpE; mRNA, messenger RNA; PFD, prefoldin.

such as geldanamycin has emerged as a promising strategy for the treatment of certain cancers<sup>55</sup>. These drugs specifically inhibit the ATPase function of HSP90. They will probably prove useful not only in cancer therapy but also in the treatment of viral diseases, owing to the fact that various pathogenic viruses hijack the HSP90 system and use it for capsid assembly<sup>56</sup>. However, the global inhibition of HSP90 is likely to result in a marked derangement of cellular circuitry, and it would be desirable to find ways to inhibit only specific aspects of HSP90 function.

### From ribosome to folded protein

The vectorial synthesis of polypeptides on the ribosome has important implications in the folding process that are only partly understood. Key questions concern the stage at which the nascent chain begins to fold and the extent to which the translation process modulates the free-energy landscape of folding. In addressing these issues, it is useful to first consider small, single-domain proteins, which tend to fold spontaneously *in vitro*. The translation process for such proteins seems to increase the risk of misfolding and aggregation considerably, because an incomplete nascent polypeptide is unable to fold into a stable native conformation<sup>57,58</sup> and the local concentration of nascent chains in the context of polyribosomes is very high. Furthermore, the exit channel of the large ribosomal subunit, which is ~100 Å long but at most 20 Å wide, is unfavourable to folding beyond  $\alpha$ -helices and small tertiary elements that may begin to form near the tunnel exit<sup>59–61</sup>; it

thus prevents the C-terminal 30–40 amino-acid residues of the chain from participating in long-range interactions that are necessary for cooperative domain folding. As a consequence, productive folding may occur only after the complete protein has emerged from the ribosome<sup>57,62</sup>. Because translation is relatively slow (~4–20 amino acids s<sup>-1</sup>), nascent chains are exposed in partially folded, aggregation-sensitive states for considerable periods of time. Moreover, non-native intrachain contacts formed during translation or interactions with the highly charged ribosomal surface could delay folding after completion of synthesis. For these reasons, nascent chains are thought to interact co-translationally with ribosome-bound chaperones, which inhibit their premature (mis)folding and maintain the nascent chain in a non-aggregated, folding-competent state (Fig. 5). For example, the bacterial trigger factor<sup>63</sup> binds to the small titin I27 chain (~120 amino acids) throughout translation<sup>64</sup>, presumably delaying chain collapse until the complete  $\beta$ -sandwich domain has emerged from the ribosome and is available for folding. Moreover, the aggregation of nascent chains is disfavoured by the densely packed, pseudohelical arrangement of ribosomes in polyribosome complexes — an organization that maximizes the distance between nascent-chain exit sites on adjacent ribosomes<sup>65</sup>.

Although single-domain proteins will reach their native state post-translationally, multidomain proteins may undergo domain-wise co-translational folding, as independently folding structural units (~50–300 amino acids in length) emerge sequentially from the



ribosome<sup>66,67</sup>. This process avoids non-native interdomain contacts, thus smoothing the folding-energy landscape for large proteins<sup>66,68</sup>. Sequential domain folding during translation, which is highly efficient on eukaryotic ribosomes, probably promoted the explosive evolution of complex multidomain proteins in eukaryotes<sup>66,68</sup>. Co-translational folding is thought to be aided by the slower elongation speed of eukaryotic ribosomes ( $\sim 4$  amino acids  $s^{-1}$  in eukaryotes versus  $\sim 20$  amino acids  $s^{-1}$  in bacteria) and as a result of various adaptations of the folding machinery. For example, eukaryotic ribosomes bind specialized HSP70 chaperone complexes (Fig. 5) and the binding and release of the canonical HSC70 from nascent chains may be coordinated with translation speed so as to support domain-wise folding. The eukaryotic chaperonin TRiC is recruited to nascent chains by HSC70 (ref. 69) and other upstream factors, such as prefoldin<sup>31</sup>, allowing co-translational folding. Moreover, fine-tuning of co-translational folding may be achieved by translational pausing at rare codons<sup>70</sup>. Overall, the eukaryotic translation and chaperone machinery has been highly optimized through evolution, ensuring efficient folding for the bulk of newly synthesized proteins<sup>71</sup>.

The chaperone pathways operating in the endoplasmic reticulum (ER) follow analogous organizational principles, but specialized machinery is used in disulphide-bond formation and the glycosylation of many secretory proteins<sup>72</sup>.

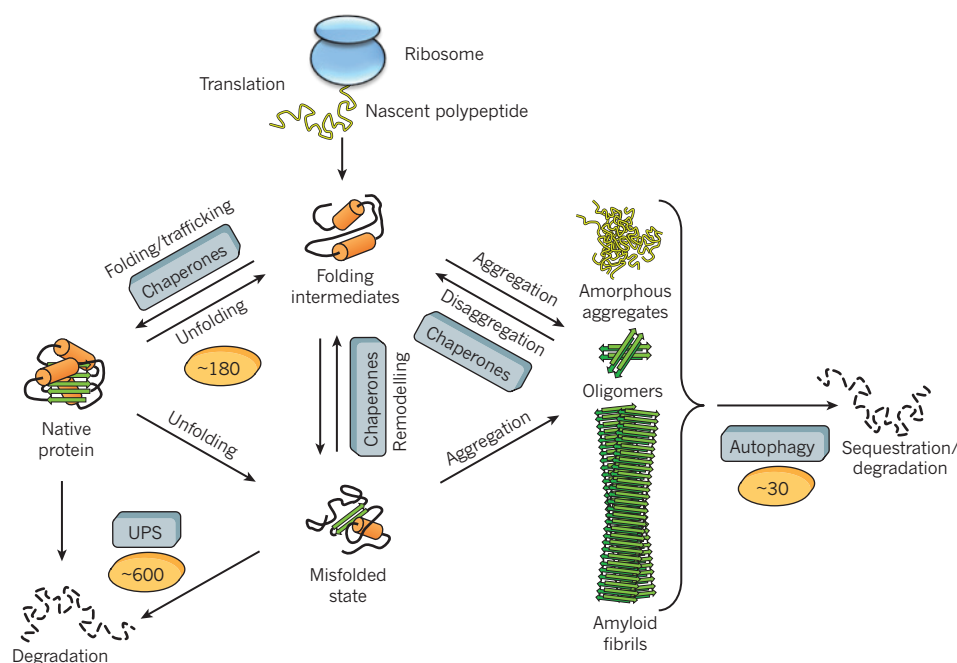
### Proteome maintenance and the proteostasis network

Although it is generally accepted that the chaperone machinery is required for initial protein folding, we are only beginning to appreciate the extent to which many proteins depend on macromolecular assistance throughout their cellular lifetime to maintain or regain their functionally active conformations. Compared with prokaryotes, the proteomes of eukaryotic cells are highly complex, comprising a much greater number and diversity of multidomain proteins. In the dynamic cellular environment, these proteins constantly face numerous challenges to their folded states; these result from post-translational modifications (phosphorylation and acetylation), changes in cell physiology and alterations in the composition and concentration of small-molecule ligands that may influence protein stability<sup>4</sup>. Moreover, 20–30% of all proteins in mammalian cells are intrinsically unstructured<sup>3</sup>; that is, they may adopt defined three-dimensional conformations only after binding to other macromolecules or membrane surfaces. Such proteins probably require assistance to avoid aberrant interactions and aggregation,

particularly when their concentration is increased and they are not in complexes with partner molecules<sup>73</sup>.

These considerations help to explain why cells must invest in an extensive network of factors, comprising  $\sim 800$  proteins in human cells ( $\sim 200$  chaperones and co-chaperones and  $\sim 600$  UPS and autophagy components), which cooperate to maintain the conformational integrity of the proteome and provide adaptation to changes in the environment. This proteostasis network integrates general and specialized chaperone components for proper protein folding and trafficking with the machinery for disaggregation and proteolytic degradation of irreversibly misfolded proteins (the UPS and the autophagy system) (Fig. 6). The remarkable complexity of the system arises from the expansion, in multicellular organisms, of the diversity of regulatory components for the major chaperone systems (HSP70 and HSP90)<sup>26</sup> and of factors functionally coupling these chaperones with the UPS and the autophagy system<sup>27,74,75</sup>. For example, various HSP70 cofactors, such as the BCL2-associated athanogene (BAG) family proteins and certain HSP40s, contain ubiquitin-like or ubiquitin-interacting domains<sup>74</sup>. The HSP70 and HSP90 cofactor known as carboxyl terminus of Hsp70-interacting protein (CHIP) has E3 ubiquitin ligase activity and channels certain mutant or damaged proteins towards proteasomal degradation<sup>74</sup>. Notably, CHIP is only one of several hundred different E3 ligases, which reflects the enormous importance of proteolytic pathways for proteostasis and cell regulation. Interestingly, whereas the clearance of misfolded protein species by the UPS requires that these molecules are maintained in a non-aggregated state by chaperones, disposal by autophagy is thought to involve active mechanisms to force such molecules into larger, presumably less toxic, aggregates<sup>76,77</sup>. These inclusions are often deposited at specific subcellular sites close to the microtubule-organizing centre, referred to as the aggresome<sup>78</sup>.

The proteostasis network is regulated by several interconnected signalling pathways, some of which are stress responsive and ensure that cellular protein folding and/or degradation is adapted to avoid the accumulation of misfolded and aggregation-prone species (Fig. 6). These pathways include the cytosolic stress response and the unfolded protein response of the ER and mitochondria, as well as signalling pathways that control ribosome biogenesis and translational capacity (Box 1). How the inputs from these different branches are coordinated and fine-tuned is only partly understood, but proteostasis capacity and responsiveness to stress may vary considerably in different cell types<sup>79</sup>.



**Figure 6 | Protein fates in the proteostasis network.** The proteostasis network integrates chaperone pathways for the folding of newly synthesized proteins, for the remodelling of misfolded states and for disaggregation with the protein degradation mediated by the UPS and the autophagy system. Approximately 180 different chaperone components and their regulators orchestrate these processes in mammalian cells, whereas the UPS comprises  $\sim 600$  and the autophagy system  $\sim 30$  different components. The primary effort of the chaperone system is in preventing aggregation, but machinery for the disaggregation of aggregated proteins has been described in bacteria and fungi, involving oligomeric AAA<sup>+</sup> proteins such as HSP104 and the *E. coli* molecular chaperone protein ClpB, which cooperate with HSP70 chaperones<sup>25</sup>. A similar activity has been detected in metazoans, but the components involved have not yet been defined<sup>83</sup>.

## BOX 1

# Signalling pathways in proteostasis and ageing

The expression of stress-inducible chaperone proteins (such as HSP70, HSP40, HSP90 and small HSPs) in the cytosol is governed by the heat-shock response<sup>96</sup>. The genes encoding these proteins are transcriptionally regulated by the HSF-1 and FOXO (DAF-16 in *C. elegans*) transcription factors.

The unfolded protein response (UPR)<sup>72</sup> of the ER adjusts the folding capacity of the secretory pathway by upregulating ER chaperones and/or attenuating protein synthesis by means of the transcription factors IRE1, PERK and ATF6.

The mitochondrial UPR<sup>97,98</sup> is activated by conformational stress in mitochondria and increases resistance to oxidative damage.

Ageing and longevity pathways are coupled to the regulation of stress-protective pathways<sup>5,99</sup>. Specifically, the upregulation of stress-protection factors such as chaperones by HSF-1 and FOXO is required for the lifespan-extending effect of mutations in the insulin and insulin-like growth factor I (IGF-I) receptor pathway. Autophagy, a process required for the recycling of organelles and the removal of large protein aggregates, is also necessary for lifespan extension and youthfulness in *C. elegans*. Autophagy is downregulated by the mammalian target of rapamycin (TOR) kinase when nutrients are plentiful<sup>99</sup> and is upregulated by FOXO<sup>81</sup>. Dietary restriction, which extends lifespan in model organisms, is also coupled with HSF-1 and FOXO activation<sup>81,100</sup>.

## Proteostasis collapse in ageing and disease

The accumulation of misfolded and/or oxidized proteins in cells during ageing is a challenge to the proteostasis system and eventually results in the deposition of aggregates, as shown in model organisms such as *Caenorhabditis elegans* and *Drosophila*<sup>80,81</sup>. The inability of cells to restore normal proteostasis may result in disease, and even in cell death. Indeed, numerous diseases are now recognized to be associated with aberrant protein folding and are usually categorized as loss-of-function or toxic gain-of-function diseases, although specific pathological states often show elements of both groups. The former are generally caused by inherited mutations and include numerous disorders such as cystic fibrosis, lysosomal storage diseases and  $\alpha$ 1-antitrypsin deficiency. The latter, gain-of-function disorders, include type 2 diabetes and the major neurodegenerative conditions (Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis and Alzheimer's disease) and are either sporadic or caused by mutations that render specific proteins more aggregation prone. These gain-of-function diseases are typically age related and are caused by the accumulation of amyloid or amyloid-like aggregates of the disease protein. A plausible explanation for the late onset of these diseases is provided by recent evidence from model organisms that the signalling pathways that regulate proteostasis are integrated with the genetic and epigenetic pathways that control longevity<sup>82,83</sup> (Box 1). Thus, the age-related decline in proteostasis and specifically in the inability to upregulate chaperones in response to conformational stresses would trigger disease manifestation and, in turn, accelerate proteostasis collapse<sup>81,84,85</sup>.

Although the toxic principle operating in these disorders is far from understood, a consensus is emerging that soluble oligomeric aggregates, which may be 'on-pathway' or 'off-pathway' towards fibril formation, are the primary cytotoxic species<sup>16,18</sup> (Fig. 1). One prominent hypothesis suggests that these oligomers expose promiscuous hydrophobic surfaces that can mediate aberrant interactions with several other proteins or with cellular membranes<sup>16,17</sup>. In support of this proposal, a recent proteomics study in human cells showed that certain metastable proteins are targeted preferentially by such interactions, resulting in their co-aggregation with the amyloidogenic disease protein<sup>86</sup>. The co-aggregating proteins are generally large in size and are enriched in intrinsically unstructured regions, properties that are coupled with a high degree of functionality. Accordingly, they tend to occupy essential hub positions in cellular protein networks, including transcriptional regulation, translation and maintenance of cell architecture, suggesting that their sequestration by the amyloid aggregates results in multifactorial toxicity. An interesting manifestation of this toxicity mechanism is the recent demonstration that aggregating mutant p53 may exert dominant oncogenic potential by sequestering wild-type p53 into co-aggregates, resulting in a complete loss of p53 function<sup>87</sup>.

Aggregate toxicity may be exacerbated by the inability of affected cells to adequately respond to stress stimuli<sup>86</sup>. This is consistent with

recent evidence that aberrantly folded protein species may interfere with central proteostasis functions, including protein folding and clearance mechanisms<sup>88,89</sup>. Notably, the overexpression of members of the HSP70 system has been shown to inhibit the formation of toxic oligomers and to prevent the formation of amyloid aggregates for different disease proteins<sup>24,90,91</sup>. In the case of polyglutamine-repeat proteins, which cause Huntington's disease and several related neurodegenerative disorders, HSP70 cooperates with the chaperonin TRiC to prevent the accumulation of potentially toxic oligomers<sup>47</sup>, which is reminiscent of the functional cooperation between these chaperone systems in *de novo* protein folding.

On the basis of these findings, the pharmacological upregulation of chaperone function promises to open up new strategies for the treatment of numerous pathological states associated with aberrant folding and aggregation. Proof-of-principle experiments using small-molecule compounds to increase chaperone synthesis and rebalance proteostasis (for example, by activating heat-shock transcription factor-1 (HSF-1)-regulated pathways) have already demonstrated efficacy in loss-of-function and toxic gain-of-function disease models<sup>5,6,92,93</sup>. Likewise, recently identified proteasome activators<sup>94</sup> have the potential to accelerate the clearance of toxic protein species, particularly when applied in combination with chaperone upregulation. Unlike conventional drugs, such 'proteostasis regulators' would not be disease-specific or protein-specific, and thus may be applicable to a whole group of related diseases — a new concept in medical practice.

## Outlook

Studies over the past two decades have provided fascinating insight into the mechanics of chaperone-assisted protein folding, but there are still major gaps in our understanding of how the pathways of folding in the cell differ from those studied in the test tube. Progress is being held back by the problem that the sophisticated biophysical methods used to characterize folding intermediates *in vitro* are not easily transferable to the *in vivo* situation. Major innovation potential can thus be expected from the development of advanced imaging techniques, eventually allowing us to monitor conformational changes in a single polypeptide chain as it emerges from the ribosome, performs its biological function and is finally degraded in the living cell. Much research will also be stimulated by the emerging concept that molecular chaperones function as the central element of a much larger cellular network of proteostasis control, comprising, in addition, the protein biogenesis machinery as well as the UPS and the autophagy system. Unravelling the complex regulatory circuitry of this network and understanding why it loses its grip during ageing will pose a major challenge for years to come. Solving this problem will require a broad systems-biology approach relying on a combination of ribosome profiling, quantitative proteomics and computational modelling. How cells react to conformational stress or proteostasis deficiency at the proteome level is unclear. Key questions include determining how certain aberrantly folding proteins aggregate



into toxic species whereas others are degraded, how the composition of the proteosome changes during ageing, what the signature of a youthful proteome is, and how we can find ways to maintain it for longer as we age. Addressing these and related issues not only offers great opportunities for intervention with numerous, currently incurable diseases but will also eventually reveal the fundamentally important relationship between proteostasis and longevity. ■

1. Dobson, C. M., Sali, A. & Karplus, M. Protein folding — a perspective from theory and experiment. *Angew. Chem. Int. Edn Engl.* **37**, 868–893 (1998).
2. Bartlett, A. I. & Radford, S. E. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nature Struct. Mol. Biol.* **16**, 582–588 (2009).
3. Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18**, 756–764 (2008).
4. Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78**, 959–991 (2009).
5. Morimoto, R. I. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev.* **22**, 1427–1438 (2008).
6. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
7. Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571–580 (1996).
8. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004).
9. Herbst, R., Schafer, U. & Seckler, R. Equilibrium intermediates in the reversible unfolding of firefly (*Photinus pyralis*) luciferase. *J. Biol. Chem.* **272**, 7099–7105 (1997).
10. Ellis, R. J. & Minton, A. P. Protein aggregation in crowded environments. *Biol. Chem.* **387**, 485–497 (2006).
11. Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–671 (2009).
12. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- This seminal study puts forward the idea that chaperones function in buffering the otherwise deleterious consequences of mutations.**
13. Skach, W. R. Cellular mechanisms of membrane protein folding. *Nature Struct. Mol. Biol.* **16**, 606–612 (2009).
14. Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
15. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41**, 161–172 (2011).
- This exciting paper describes, at atomic resolution, the structural features of a non-native folding intermediate that are critical for amyloidogenic aggregation.**
16. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
17. Bolognesi, B. *et al.* ANS binding reveals common features of cytotoxic amyloid species. *ACS Chem. Biol.* **5**, 735–740 (2010).
- This paper provides evidence that the exposure of hydrophobic surfaces by oligomeric aggregation intermediates correlates with their toxicity.**
18. Kaye, R. *et al.* Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486–489 (2003).
19. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Struct. Mol. Biol.* **16**, 574–581 (2009).
20. Langer, T. *et al.* Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**, 683–689 (1992).
21. Frydman, J., Nimmesgern, E., Ohtsuka, K. & Hartl, F. U. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370**, 111–117 (1994).
22. Ellis, R. J. Molecular chaperones: assisting assembly in addition to folding. *Trends Biochem. Sci.* **31**, 395–401 (2006).
23. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
24. Auluck, P. K., Chan, H. Y. E., Trojanowski, J. Q., Lee, V. M. Y. & Bonini, N. M. Chaperone suppression of  $\alpha$ -synuclein toxicity in a *Drosophila* model for Parkinson's disease. *Science* **295**, 865–868 (2002).
25. Mayer, M. P. Gymnastics of molecular chaperones. *Mol. Cell* **39**, 321–331 (2010).
26. Kampina, H. H. & Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Rev. Mol. Cell Biol.* **11**, 579–592 (2010).
27. Arndt, V. *et al.* Chaperone-assisted selective autophagy is essential for muscle maintenance. *Curr. Biol.* **20**, 143–148 (2010).
28. Rüdiger, S., Buchberger, A. & Bukau, B. Interaction of Hsp70 chaperones with substrates. *Nature Struct. Biol.* **4**, 342–349 (1997).
- A key paper in understanding how chaperones recognize their substrates.**
29. Rousseau, F., Serrano, L. & Schymkowitz, J. W. H. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.* **355**, 1037–1047 (2006).
30. Sharma, S. K., De los Rios, P., Christen, P., Lustig, A. & Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nature Chem. Biol.* **6**, 914–920 (2010).
31. Frydman, J. Folding of newly translated proteins *in vivo*: the role of molecular chaperones. *Annu. Rev. Biochem.* **70**, 603–647 (2001).
32. Horwich, A. L. & Fenton, W. A. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q. Rev. Biophys.* **42**, 83–116 (2009).
33. Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T. & Taguchi, H. A systematic survey of *in vivo* obligate chaperonin-dependent substrates. *EMBO J.* **29**, 1552–1564 (2010).
34. Raineri, E., Ribeca, P., Serrano, L. & Maier, T. A more precise characterization of chaperonin substrates. *Bioinformatics* **26**, 1685–1689 (2010).
35. Tartaglia, G. G., Dobson, C. M., Hartl, F. U. & Vendruscolo, M. Physicochemical determinants of chaperone requirements. *J. Mol. Biol.* **400**, 579–588 (2010).
36. Xu, Z. H., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–(ADP)<sub>7</sub> chaperonin complex. *Nature* **388**, 741–749 (1997).
37. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
38. Tang, Y. C. *et al.* Structural features of the GroEL–GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**, 903–914 (2006).
39. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**, 112–122 (2010).
40. Thirumalai, D. & Lorimer, G. H. Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269 (2001).
41. Lin, Z., Madan, D. & Rye, H. S. GroEL stimulates protein folding through forced unfolding. *Nature Struct. Mol. Biol.* **15**, 303–311 (2008).
42. Sharma, S. *et al.* Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* **133**, 142–153 (2008).
43. Munoz, I. G. *et al.* Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nature Struct. Mol. Biol.* **18**, 14–19 (2011).
44. Douglas, N. R. *et al.* Dual action of ATP hydrolysis couples lid closure to substrate release into the Group II chaperonin chamber. *Cell* **144**, 240–252 (2011).
45. Reissmann, S., Parnot, C., Booth, C. R., Chiu, W. & Frydman, J. Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nature Struct. Mol. Biol.* **14**, 432–440 (2007).
46. Kitamura, A. *et al.* Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. *Nature Cell Biol.* **8**, 1163–1170 (2006).
47. Behrends, C. *et al.* Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. *Mol. Cell* **23**, 887–897 (2006).
48. Tam, S., Geller, R., Spiess, C. & Frydman, J. The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nature Cell Biol.* **8**, 1155–1162 (2006).
49. Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Rev. Mol. Cell Biol.* **11**, 515–528 (2010).
50. McClellan, A. J. *et al.* Diverse cellular functions of the Hsp90 molecular chaperone uncovered using systems approaches. *Cell* **131**, 121–135 (2007).
51. Scheufler, C. *et al.* Structure of TPR domain–peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multichaperone machine. *Cell* **101**, 199–210 (2000).
52. Wandinger, S. K., Richter, K. & Buchner, J. The Hsp90 chaperone machinery. *J. Biol. Chem.* **283**, 18473–18477 (2008).
53. Ali, M. M. U. *et al.* Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* **440**, 1013–1017 (2006).
54. Shiau, A. K., Harris, S. F., Southworth, D. R. & Agard, D. A. Structural analysis of *E. coli* hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell* **127**, 329–340 (2006).
55. Neckers, L. Heat shock protein 90: the cancer chaperone. *J. Biosci.* **32**, 517–530 (2007).
56. Geller, R., Vignuzzi, M., Andino, R. & Frydman, J. Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. *Genes Dev.* **21**, 195–205 (2007).
- This seminal study describes the requirement of HSP90 in viral assembly, outlining a strategy for antiviral treatment based on HSP90 inhibition.**
57. Eichmann, C., Preissler, S., Riek, R. & Deuring, E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **107**, 9111–9116 (2010).
58. Cabrita, L. D., Hsu, S. T., Launay, H., Dobson, C. M. & Christodoulou, J. Probing ribosome–nascent chain complexes produced *in vivo* by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **106**, 22239–22244 (2009).
59. Lu, J. L. & Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nature Struct. Mol. Biol.* **12**, 1123–1129 (2005).
60. Woolhead, C. A., McCormick, P. J. & Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725–736 (2004).
61. O'Brien, E. P., Hsu, S.-T. D., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Transient tertiary structure formation within the ribosome exit port. *J. Am. Chem. Soc.* **132**, 16928–16937 (2010).
62. Elcock, A. H. Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput. Biol.* **2**, e98 (2006).
63. Ferbitz, L. *et al.* Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).
64. Kaiser, C. M. *et al.* Real-time observation of trigger factor function on translating ribosomes. *Nature* **444**, 455–460 (2006).
65. Brandt, F. *et al.* The native 3D organization of bacterial polysomes. *Cell* **136**, 261–271 (2009).

66. Netzer, W. J. & Hartl, F. U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343–349 (1997).
  67. Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F. U. Co-translational domain folding as the structural basis for the rapid *de novo* folding of firefly luciferase. *Nature Struct. Biol.* **6**, 697–705 (1999).
  68. Agashe, V. R. *et al.* Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* **117**, 199–209 (2004).
  69. Cuellar, J. *et al.* The structure of CCT-Hsc70<sub>NBD</sub> suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nature Struct. Mol. Biol.* **15**, 858–864 (2008).
  70. Zhang, G. & Ignatova, Z. Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS ONE* **4**, e5036 (2009).
  71. Vabulas, R. M. & Hartl, F. U. Protein synthesis upon acute nutrient restriction relies on proteasome function. *Science* **310**, 1960–1963 (2005).
  72. Buchberger, A., Bukau, B. & Sommer, T. Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* **40**, 238–252 (2010).
  73. Vavouri, T., Sempke, J. I., Garcia-Verdugo, R. & Lehner, B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **138**, 198–208 (2009).
  74. Arndt, V., Rogon, C. & Hohfeld, J. To be, or not to be — molecular chaperones in protein degradation. *Cell. Mol. Life Sci.* **64**, 2525–2541 (2007).
  75. Gamerding, M. *et al.* Protein quality control during aging involves recruitment of the macroautophagy pathway by BAG3. *EMBO J.* **28**, 889–901 (2009).
  76. Kaganovich, D., Kopito, R. & Frydman, J. Misfolded proteins partition between two distinct quality control compartments. *Nature* **454**, 1088–1095 (2008).
  77. Iwata, A., Riley, B. E., Johnston, J. A. & Kopito, R. R. HDAC6 and microtubules are required for autophagic degradation of aggregated huntingtin. *J. Biol. Chem.* **280**, 40282–40292 (2005).
  78. Kopito, R. R. Aggresomes, inclusion bodies and protein aggregation. *Trends Cell Biol.* **10**, 524–530 (2000).
  79. Kern, A., Ackermann, B., Clement, A. M., Duerk, H. & Behl, C. HSF1-controlled and age-associated chaperone capacity in neurons and muscle cells of *C. elegans*. *PLoS ONE* **5**, e8568 (2010).
  80. David, D. C. *et al.* Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol.* **8**, e1000450 (2010).
  81. Demontis, F. & Perrimon, N. FOXO/4E-BP signaling in *Drosophila* muscles regulates organism-wide proteostasis during aging. *Cell* **143**, 813–825 (2010).
  82. Morley, J. F. & Morimoto, R. I. Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* **15**, 657–664 (2004).
- This pioneering study provides important insight into the relationship between molecular chaperone functions and longevity.**
83. Cohen, E., Bieschke, J., Perciavalle, R. M., Kelly, J. W. & Dillin, A. Opposing activities protect against age-onset proteotoxicity. *Science* **313**, 1604–1610 (2006).
- An exciting study demonstrating that active disaggregation and the forced formation of large inclusions prevent the accumulation of toxic aggregate species in *C. elegans*.**
84. Ben-Zvi, A., Miller, E. A. & Morimoto, R. I. Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proc. Natl Acad. Sci. USA* **106**, 14914–14919 (2009).
  85. Cohen, E. *et al.* Reduced IGF-1 signaling delays age-associated proteotoxicity in mice. *Cell* **139**, 1157–1169 (2009).
  86. Olzscha, H. *et al.* Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67–78 (2011).
- This paper demonstrates the existence of a metastable sub-proteome that is at risk of co-aggregating with amyloid-forming disease proteins.**
87. Xu, J. *et al.* Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. *Nature Chem. Biol.* **7**, 285–295 (2011).
- This interesting study expands the range of diseases promoted by proteostasis deficiency to cancer.**
88. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin–proteasome system by protein aggregation. *Science* **292**, 1552–1555 (2001).
- This key paper demonstrates that protein aggregation can interfere with protein degradation.**
89. Gidalevitz, T., Ben-Zvi, A., Ho, K. H., Brignull, H. R. & Morimoto, R. I. Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* **311**, 1471–1474 (2006).
  90. Schaffar, G. *et al.* Cellular toxicity of polyglutamine expansion proteins: mechanism of transcription factor deactivation. *Mol. Cell* **15**, 95–105 (2004).
  91. Lotz, G. P. *et al.* Hsp70 and Hsp40 functionally interact with soluble mutant huntingtin oligomers in a classic ATP-dependent reaction cycle. *J. Biol. Chem.* **285**, 38183–38193 (2010).
  92. Sittler, A. *et al.* Geldanamycin activates a heat shock response and inhibits huntingtin aggregation in a cell culture model of Huntington's disease. *Hum. Mol. Genet.* **10**, 1307–1315 (2001).
  93. Mu, T. W. *et al.* Chemical and biological approaches synergize to ameliorate protein-folding diseases. *Cell* **134**, 769–781 (2008).
  94. Lee, B.-H. *et al.* Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* **467**, 179–184 (2010).
- This study describes the first drug-like molecule that can activate proteasome function, thus providing a means to enhance the clearance of aberrantly folded proteins.**
95. Jahn, T. R. & Radford, S. E. The Yin and Yang of protein folding. *FEBS J.* **272**, 5962–5970 (2005).
  96. Vabulas, R. M., Raychaudhuri, S., Hayer-Hartl, M. & Hartl, F. U. Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harb. Perspect. Biol.* **2**, a004390 (2010).
  97. Ryan, M. T. & Hoogenraad, N. J. Mitochondrial-nuclear communications. *Annu. Rev. Biochem.* **76**, 701–722 (2007).
  98. Haynes, C. M. & Ron, D. The mitochondrial UPR — protecting organelle protein homeostasis. *J. Cell Sci.* **123**, 3849–3855 (2010).
  99. Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
  100. Westerheide, S. D., Ankar, J., Stevens, S. M., Jr, Sistonen, L. & Morimoto, R. I. Stress-inducible regulation of heat shock factor 1 by the deacetylase SIRT1. *Science* **323**, 1063–1066 (2009).

**Acknowledgements** We thank W. Balch, A. Dillin, J. Kelly, R. Morimoto and P. Reinhart for discussions about proteostasis, and thank the members of our laboratory for comments on the manuscript. We apologize to all those whose important work could not be cited owing to space limitations.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to F.U.H. ([uhartl@biochem.mpg.de](mailto:uhartl@biochem.mpg.de)).



# Nuclear export dynamics of RNA–protein complexes

David Grünwald<sup>1</sup>, Robert H. Singer<sup>2</sup> & Michael Rout<sup>3</sup>

**The central dogma of molecular biology — DNA makes RNA makes proteins — is a flow of information that in eukaryotes encounters a physical barrier: the nuclear envelope, which encapsulates, organizes and protects the genome. Nuclear-pore complexes, embedded in the nuclear envelope, regulate the passage of molecules to and from the nucleus, including the poorly understood process of the export of RNAs from the nucleus. Recent imaging approaches focusing on single molecules have provided unexpected insight into this crucial step in the information flow. This review addresses the latest studies of RNA export and presents some models for how this complex process may work.**

Since its first description in electron micrographs<sup>1</sup>, our understanding of the nuclear-pore complex (NPC), arguably the largest nanomachine in the cell, has increased steadily. We are now at the point where we have a comprehensive overview of the NPC components and their contribution to its structure, as well as initial insights into the mechanism of NPC assembly and a sound understanding of the principal functions of the NPC<sup>2</sup>. The 100-nm diameter NPC has a core structure consisting of a hollow cylinder embedded in the nuclear envelope, which displays an eight-fold symmetry of about 30 different proteins termed nucleoporins (Nups). The NPC acts as the gateway between the nucleus and the cytoplasm; only those macromolecules carrying specific import and export signals are permitted to pass through the central channel of the NPC, although water and metabolites can pass through freely<sup>3,4</sup>. The NPC consists of several major domains (Fig. 1): the selective central channel, or central transporter region; the core scaffold that supports the central channel; the transmembrane regions; the nuclear basket; and the cytoplasmic filaments<sup>5</sup>. The central channel is filled and surrounded with a distinct class of Nup that has numerous large domains rich in phenylalanine and glycine repeats, termed FG Nups. It is this central channel and the FG Nups that seem sufficient to mediate selective receptor-mediated transport<sup>6,7</sup>. The nuclear basket consists of eight filaments that reach into the nucleoplasm, attached to each other by a ring at the end. Electron microscopy tomographs have shown that filaments extend from this basket into the nucleus<sup>8,9</sup>. The cytoplasmic filaments are less ordered, forming highly mobile molecular rods projecting into the cytoplasm. The reach of NPCs can extend about 100 nm into the nucleus and cytoplasm<sup>10,11</sup>.

The transport of molecules through the NPC is restricted by size; below a mass of approximately 60 kDa, macromolecules can passively diffuse across the NPC (albeit slowly, as the molecule approaches the 60 kDa cut-off<sup>12</sup>). The exact cut-off size remains unclear, although several studies have addressed this issue using various sized molecular probes<sup>13–15</sup>. Moreover, even small macromolecules (that is, below this cut-off) also frequently contain a nuclear localization signal that allows usage of the receptor-mediated transport pathways<sup>16</sup>. Hence, to be shipped as cargoes across the NPC, transport signals seem mandatory for almost all macromolecules: nuclear localization sequences (NLSs) for import into the nucleus and nuclear export sequences (NESs) for export. These signals are recognized by transport factors, each with specific signal preferences. Many transport receptors belong to the

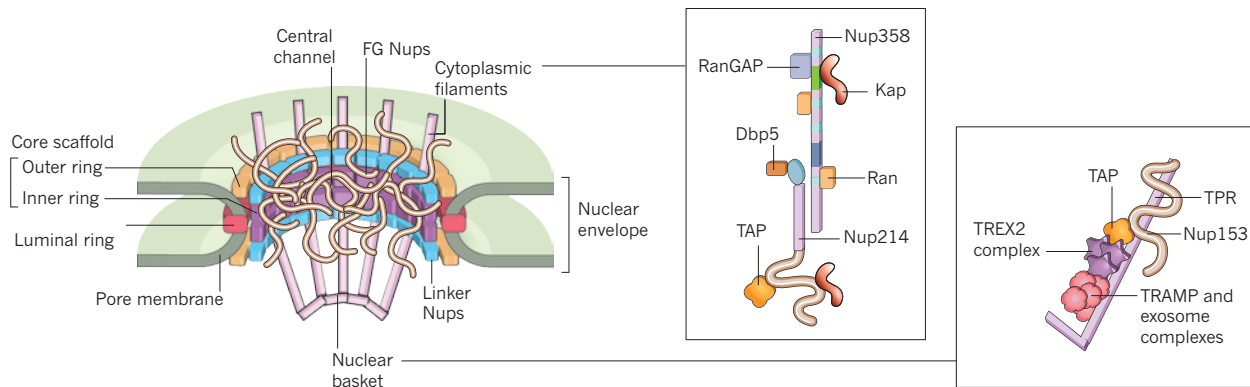
karyopherin (importin and exportin) families, characterized by a shared  $\alpha$ -superhelical structure. Karyopherins can bind to the NLSs or NESs of their cognate cargoes, to the FG Nups and to the GTPase Ran<sup>17</sup>. For NLS-containing proteins, an import cycle starts with the formation of the cargo–karyopherin complex in the cytoplasm, which seems to be the rate-limiting step *in vivo*<sup>18,19</sup>, and then proceeds with translocation through the NPC and, finally, disassembly of the complex on the nuclear side by the binding of Ran-GTP to the karyopherin<sup>3,4,17</sup>. This process is driven by a Ran-GTP gradient across the nuclear envelope; Ran cofactors localized to the nucleoplasm and cytoplasm and a Ran-specific nuclear transport factor (NTF2) maintain a high concentration of nuclear Ran-GTP and of cytoplasmic Ran-GDP<sup>3,16</sup>. Protein export has been shown to be governed by very similar principles to the well-studied import machinery. An NES on a cargo is recognized by a cognate karyopherin–Ran-GTP dimer in the nucleus and, after translocation across the NPC, the NES-cargo–karyopherin–Ran-GTP complex is disassembled on the cytoplasmic side, through activation of Ran GTPase activity by cytoplasmic RanGAP, achieving directionality<sup>20,21</sup>. As we discuss below, not all transport factors require Ran, nor belong to the karyopherin family; however, notably, all transport receptors can interact directly with FG Nups<sup>22</sup>.

An open question is how transport selectivity is achieved by the available components of the NPC. It is clear that FG Nups are essential *in toto*, not surprisingly given that they are the docking sites of the complex for transport factors. Deletions of individual FG repeat domains in yeast are not overtly harmful; however, various combinations of these deletions are, and there is a critical mass of deletions above which the NPC cannot function<sup>6</sup>. Numerous lines of evidence show that the FG repeat domains are natively unfolded<sup>23,24</sup>, and they form a tangle of filaments needed to establish the transport barrier in the central channel of the NPC. Reagents that disrupt this tangle also disrupt transport<sup>25–27</sup>.

## Models of nucleocytoplasmic transport

Although current models explaining the molecular mechanism of selective nuclear transport differ in their details, they agree that the FG repeat domains in the central channel of the NPC form a dense and dynamic network of filaments that blocks translocation of inert molecules, and that this barrier is overcome with the help of transport receptors<sup>18,25,28–31</sup> (Fig. 2). A common idea in these various models is that the FG repeat domains conspire to produce an unfavourable

<sup>1</sup>Delft University of Technology, Faculty of Applied Sciences, Kavli Institute of NanoScience, Department of Bionanoscience, 2628 CJ Delft, The Netherlands. <sup>2</sup>Department of Anatomy and Structural Biology, Gruss Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, New York 10461, USA. <sup>3</sup>The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.



**Figure 1 | Nuclear-pore complex basic structure and function.** A schematic representation of the NPC. Major structural elements are indicated. The cytoplasmic and nuclear extensions of the vertebrate NPC's periphery are indicated on the cytoplasmic surface as Nup214 and Nup358, which carry

factors that aid the egress of cargo such as ribonucleoproteins (RNPs) from the NPC, and on the nuclear surface as TPR (translocated promoter region), the nuclear-basket filament protein that carries factors aiding late RNP processing steps and the first stages of RNP export. See text for more details.

environment for diffusion of inert molecules through the NPC's central channel. This barrier is overcome for cargoes with cognate transport receptors that bind to the FG repeats, thus counteracting the exclusion. In a sense, the NPC can be considered an enzyme for transport, in which only the correct substrates (such as transport factors and their cargoes) can bind to the active site and so pass across the nuclear envelope<sup>28,29</sup>. The directionality of transport is intimately linked to the release of cargo from the transport complex being allowed only on the correct side of the NPC<sup>24–29,31–34</sup>.

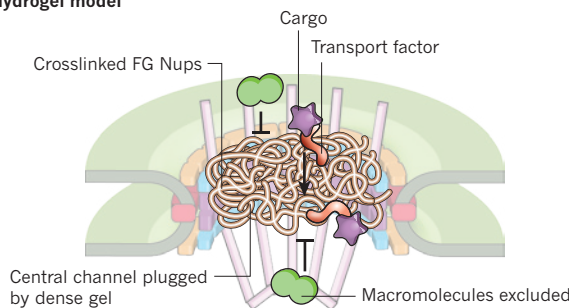
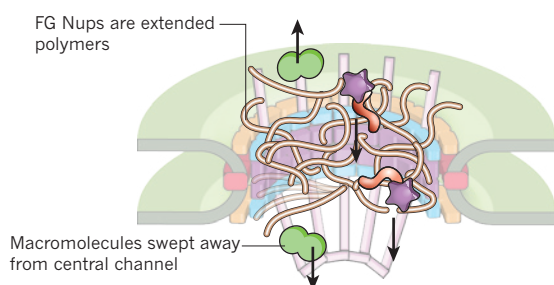
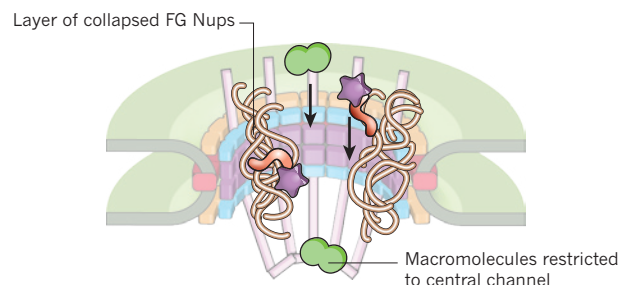
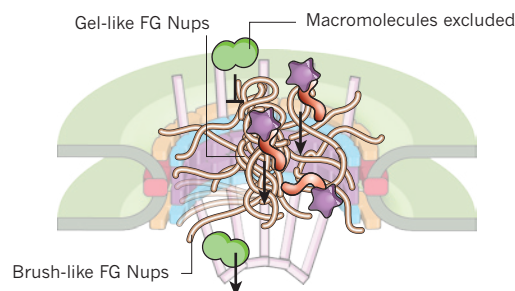
### Transport and single-molecule microscopy

An understanding of the precise steps that are involved in crossing the NPC is still missing. However, emerging single-molecule imaging approaches are showing the real-time dynamics of nuclear transport, and are illuminating its mechanism. Examples of these technologies are 4-Pi microscopy<sup>35,36</sup>, single point edge excitation subdiffraction microscopy<sup>37</sup>, fluorescence correlation spectroscopy (FCS)<sup>21,38</sup>, single-molecule tracking<sup>10,19,39,40</sup> and super-registration microscopy<sup>11</sup> (Box 1). The application of such approaches to determine the distribution of Nups and transport-factor-binding sites supports the notion that the NPC functionally extends far into both compartments (the nucleoplasm and cytoplasm) on either side of itself<sup>8,11,19</sup>. This agrees well with data using colloidal-gold-labelled transport cargoes and electron microscopy, which showed the cargoes docking to filaments extending dozens of nanometres from the NPC<sup>41–43</sup>. Dwell times of transport factors at the NPC have been found to range from 5 to 20 ms (Table 1). Variations in the transport factor, cargo and Ran-GTP concentration have a profound effect on the translocation times of proteins. The dwell time of the karyopherin importin- $\beta$ 1 could be reduced to 1 ms after increased concentrations of unlabelled importin- $\beta$ 1 in the cytoplasmic buffer<sup>44</sup>. In living cells, dwell times were found to be in the range of 5 to 7 ms<sup>19</sup>. Nucleocytoplasmic transport of proteins has been shown by confocal microscopy to be as high as ~1,000 molecules per NPC per second<sup>39,45</sup>. A dwell time of 5 ms translates into 200 parallel transport events per NPC per second, such that as many as 100 copies of importin- $\beta$ 1 occupy each NPC at any one time<sup>46</sup>. Notably, the presence of cargo also has an effect on dwell times by shortening the translocation process<sup>10,44</sup>, suggesting that the NPC needs to be viewed as a crowded environment. The central channel of the NPC is presumably filled with disordered FG repeat domains, unloaded and cargo-loaded transport receptors, and non-specific proteins competing to enter the NPC<sup>10,33,34,44</sup>. Molecular crowding can have two effects on NPC function<sup>47,48</sup>: transport times and binding-site availability might change based on the occupation of the central channel with transport factors, cargo and non-specific competitors; and, it might affect the folding or shape of the disordered FG repeat domains<sup>33,34,47,48</sup>. This

crowding should lead to competition for space and binding sites; in this way, transport factors with or without their cargoes, binding to the FG repeats, would tend to exclude other proteins that cannot bind in the same way to the NPC. This effect and the constant presence of transport factors in the NPC noted earlier would increase the selectivity of the NPC while maintaining its high flux rate<sup>33,34</sup>. In living cells, this high transport rate is represented by several transport factors carrying many importing and exporting cargoes, including ribonucleoproteins (RNPs). Thus, it seems that it is not the rate of passage across the NPC that limits the speed at which a cell can deliver its cargo from one side of the nuclear envelope to the other; instead, it has been shown that the formation of cargo-receptor complexes is limiting for import<sup>10,18,19,39</sup>. This point will be particularly important for considering how RNPs are delivered across the nuclear envelope.

Notably, the ability to observe single-molecule translocations at the NPC allows the direct measurement of transport efficiencies. As would be expected for a diffusion-based process, only half of the attempts made by NLS cargo to pass from the cytoplasm all the way to the nucleus are successful<sup>39</sup>. Modifications of the importin- $\beta$ 1 concentration, the Ran-GTP gradient and the cargo size have been shown to shift this balance<sup>44,49</sup>. Using fluorescence resonance energy transfer between import receptors and cargo, the directionality and the release of the complex have also been visualized<sup>50</sup>. Transport complexes move by diffusion inside the NPC and thus change their direction stochastically. Hence, cargo release is necessary to impose directionality. Recent FCS data indicated that, unless the cargo is removed from the soluble pool by interaction with immobile structures, the NPC is a bidirectional exchange catalyst, which, according to Le Chatelier's principle, will ultimately establish a steady-state balance of cargo enriched on one side of the nuclear envelope over the other<sup>19,38,50</sup>. This is in agreement with the observation that transport directionality can be inverted based on the direction of the Ran-GTP gradient<sup>11,20,21,49</sup>. The spatial location of cargo-receptor dissociation remains unclear<sup>50</sup>. The distribution of Ran and an importin- $\beta$ 1 truncation with reduced binding affinity for Ran did not indicate a clear location for the release of the receptor-cargo complex<sup>19,36</sup>. For import factors and cargoes, most data indicate that the binding-site distribution along the nuclear-cytoplasmic axis of the central channel is symmetrical, with peaks only a few nanometres off centre compared with the POM121 marker signal (Table 1), although an exception is found for the export of messenger RNA<sup>11</sup> (see below). Tracing single molecules in three dimensions also showed a non-uniform spatial distribution of importin- $\beta$ 1 across the orthogonal axis of the NPC, with higher probability densities found towards the walls of the central channel<sup>37,51</sup>. These and other data suggest that different transport pathways may follow different routes across the NPC<sup>31,52,53</sup>.



**a Hydrogel model****b Polymer brush model****c Reduction in dimensionality model****d Trees and bushes model**

**Figure 2 | Modes of transport.** Various models for how the FG Nups mediate the selective barrier function of the NPC are shown. The detailed distribution of FG repeat domains is not illustrated here. **a**, FG Nups polymerize into a gel through which transport receptors pass by binding to the FG Nups and dissolving the crosslinks<sup>25</sup>. **b**, The FG repeat filaments diffuse around their tether, and other molecules are excluded from this region. Transport factors pass through by binding to the FG Nups<sup>28,29</sup>. The FG Nups might also act as a molecular brush that collapses once transport receptors have bound other molecules<sup>24</sup>. **c**, FG Nups collapse after binding by transport factors to form a

layer along the walls of the channel. This layer is impenetrable to inert molecules but permeable to transport factors<sup>52</sup>. Inert macromolecules are able to pass through the central channel only. **d**, FG Nups form two categories of disordered filaments: collapsed coils, which are gel-like; and extended coils, which are brush-like<sup>31</sup>. Transport factors can pass through both configurations, but macromolecules are excluded. An argument can also be made (not shown) that the central channel *in vivo* will always be permeated with transport receptors, loaded or unloaded with cargo, resulting in a highly crowded environment. This could have a profound influence on the physical state of the FG Nups<sup>33,34</sup>.

## Principles and players of messenger RNP export

Our picture of nuclear transport is still mainly based on import studies, owing to the difficulty of introducing labelled export substrates into the nucleus. Import cargoes are mostly proteins that have been synthesized in the cytoplasm and are needed in the nucleus. There are also proteins that, once they reach the nucleoplasm, are exported out again by karyopherins such as XPO1 (also known as CRM1). Arguably, however, most export cargoes are RNAs, usually as complexes made of RNA and proteins. The ribosomal subunits and messenger RNPs (mRNPs) are the most abundant of these export cargoes. At around 60 kDa, the average size of protein cargoes is much smaller than mRNP cargoes, which can be as large as 100 MDa<sup>54</sup>. Such extremely large cargoes present a set of unique problems for the nuclear transport machinery (Fig. 3). First, the diameter of these cargoes can considerably exceed the diameter of the NPC central channel. Thus, to pass across the NPC, the quaternary structure of very large RNPs must be remodelled. Second, these cargoes consist of heterogeneous mixes of up to hundreds of molecules of proteins, representing dozens of protein species, packaged around an individual RNA molecule, rather than a single cargo macromolecule. The assembly of the exporting mRNP particle is clearly a complex process. Moreover, the transport machinery must distinguish between immature or incorrectly packaged mRNPs and those that are ready for export<sup>55</sup>. This task is further complicated by the fact that different mRNAs must be packaged into particles with different sizes and compositions. Third, as nucleic acids are in essence extremely long threads, they can potentially experience supercoiling problems, known as tangling.

An explanation of how such cargoes are transported may require additions to the current transport models described above. For example,

using electron microscopy to visualize cargo, complexes of up to 39 nm have been shown to cross the NPC; this includes large gold particles that cannot be deformed to squeeze through the central channel. If the gold particles cannot be deformed, then the NPC itself must change shape to accommodate transport of the particles<sup>56</sup>. Another intriguing possibility is that certain NPCs are more specialized for handling the requirements of RNP export. Using immunogold labelling, NTF2 and poly(A)<sup>+</sup> mRNAs have been shown to use different sets of NPCs in each nucleus of HL-60 cells<sup>57</sup>. This discrimination may be cell-type specific, as NTF2 has been shown to label NPCs uniformly in HeLa cell nuclei<sup>58</sup>. In yeast, NPCs adjacent to the perinuclear nucleolus lack the proteins myosin-like protein 1 (Mlp1) and Mlp2, which are important for mRNP processing, hinting that mRNP export may avoid these NPCs<sup>59</sup>.

One model of choice for RNP export has been that of Balbiani ring mRNA, found in the bloodworm larvae of the midge *Chironomus*. This RNA is huge, up to 40 kilobases (kb), and is packaged into an mRNP particle some 50 nm in diameter, far too large to fit through the NPC unaltered<sup>54,60</sup>. However, classic electron microscopy studies showed that the mRNP unravelled at the nucleoplasmic face of the NPC, and then threaded through as a thin strand while crossing the NPC. These studies, plus immuno-electron microscopy data of the proteins present in the Balbiani ring mRNP at each stage of export, have led to a picture of considerable structural and compositional rearrangement of the transcript during export<sup>54,61,62</sup>. Balbiani ring mRNA seems to be exported at the 5' end first, making it necessary to postulate a step in transport that orients the mRNA correctly, before it is threaded through the NPC. Live cell data on the mobility and inner nuclear pathways of this giant RNA complex exist, but the export dynamics of this complex remain

## BOX 1

# Microscopy used for nucleocytoplasmic transport

Fluorescence microscopy has been the standard technology in the field to establish bulk kinetic rates of nuclear transport. In the past five or so years, developments in optical technology have provided the means to use fluorescence microscopy to resolve details in the spatial and kinetic functions of the NPC. This box provides a brief overview of these applications.

## Single-molecule tracking (SMT)

In this imaging approach, the diffracted signal of a single molecule can be used to determine the position of the molecule with high precision. In combination with ultrasensitive detection and imaging frame rates of a few milliseconds, individual cargoes and receptor molecules can be followed, and their interaction time with the NPC determined<sup>10,11,38</sup>. The limitations of SMT are that only one species of molecule can be resolved in any spectral channel, and mapping between spectrally resolved channels is still diffraction limited.

## Super-registration microscopy

Super registration uses a cellular fiducial marker to allow measurement of molecular interactions at the nanometre length and millisecond timescale<sup>11</sup>.

## Single point edge excitation subdiffraction (SPEED) microscopy

In the SPEED method, a highly focused confined excitation beam (similar to confocal microscopy) is combined with ultrasensitive wide-field detection<sup>37</sup>. The resultant data have high signal-to-noise ratios and can be interpreted in three dimensions using data modelling. This approach has been used to track single molecules inside the NPC with virtual three-dimensional subdiffraction resolution.

## Highly inclined and laminated optical sheet microscopy (HILO)

In this approach, an analogy to total internal reflection fluorescence (TIRF) microscopy is used to tilt the excitation beam relative to the optical axis

of the microscope<sup>46</sup>. Although TIRF is restricted to surface-bound signals within a distance of a few hundred nanometres of the cover glass, HILO can achieve adjustable penetration depth of the sample and provide improved signal-to-noise ratios in the images, allowing SMT and super-registration imaging in the nucleus.

## Fluorescence correlation spectroscopy (FCS)

The fluctuation of fluorescence in a fixed confocal excitation spot is analysed to measure diffusive dynamics of the observed molecules<sup>21,35,36</sup>. This has the advantage of being able to resolve fast dynamics, and it has been applied to study the equilibrium conditions of NPC transit at the single-molecule level.

## 4-Pi microscopy

This approach is an extension of confocal microscopy in which two objectives are placed on opposite sides of the sample, doubling the effective numerical aperture of the detection system<sup>35,36</sup>. After deconvolution of the images, this scanning technology provides excellent resolution along the optical axis of the microscope. The method has been extended to FCS and was used to study the interaction of transport receptors with NPCs. The method also yields good registration between several spectrally resolved images.

## Super-resolution microscopy

None of the various super-resolution methods — such as structured illumination microscopy, photo-activation localization microscopy, stochastic optical reconstruction microscopy and stimulated emission depletion imaging — has yet been applied to NPC functional imaging. Ultimately, the further development of these technologies, and technical advances in optics, detectors and also in the design of fluorescent reporters, will result in high-resolution kinetic data of NPC function beyond the current state of the art.

unknown<sup>60</sup>. The ability of conventional electron microscopy to capture so many examples of its transport suggests the rate of passage of the Balbiani ring RNP is relatively slow across the NPC<sup>60</sup>. By contrast, average-sized mRNAs of a few kilobases (such as  $\beta$ -actin) are exported so fast that such major quaternary structural unfolding seems unlikely, although some remodelling must occur (see below)<sup>11</sup>. Even larger mRNAs such as the dystrophin transcript (~10 kb) may require unfolding and export on timescales of only a second<sup>40</sup>, providing some perspective on the extreme that the Balbiani ring mRNA probably represents.

Protein import into the nucleus has been shown to be GTP dependent, with directionality imposed by the Ran-GTP gradient leading to dissociation of the transport complex in the nucleus<sup>16</sup>. Although Ran is involved in upstream events leading to export (such as the import of mRNA-processing and mRNA-maturation proteins), it does not seem to provide the direct driving gradient for RNA export, which seems to be ATP dependent<sup>63–68</sup>. How export directionality is ensured is also unclear<sup>69</sup>. It is likely that the host of accessory proteins tethered to the nuclear and cytoplasmic filaments of the NPC (Fig. 1) have important roles in exchanging proteins from the mRNPs as they pass through the NPC, particularly stripping away nuclear transport factors as the mRNP exits the cytoplasmic side of the NPC, and so ensuring that transport is unidirectional. RNP export starts at the nuclear basket, where the TREX2 (3' repair exonuclease 2), TRAMP (Trf4-Air2-Mtr4p polyadenylation) and exosome complexes, involved in proofreading and final assembly of the RNP before its export, are found hovering<sup>70</sup> (Fig. 1). After processing at the basket, the RNP must

then enter the realm of the FG Nups. A key player in this stage is TAP (also known as NXF1), which forms a dimer with p15 (also known as NXT1) — these are homologues of the yeast Mex67–Mtr2 heterodimer — although p15 has been shown to be dispensable for export<sup>71</sup>. These proteins form the major transport receptors for mRNPs, as they bind both the mRNP particles and FG repeats<sup>72–74</sup>. After passing through the central channel, the RNP must then encounter the filaments on the cytoplasmic face of the NPC. Here, Nup214, Nup358 and Dbp5, a DEAD-box helicase, have also been shown to be essential for mRNA export<sup>75–77</sup>. Dbp5 functions in an ATP-dependent manner and has been proposed to supply the motor activity that would provide mechanical force to reshape the mRNP, although this motor function has not yet been conclusively shown<sup>78–80</sup>. A ratchet model has also been proposed for RNA export, in which the Dbp5-mediated removal of TAP–p15 leads to transport directionality<sup>69</sup>. Although remodelling events could be used to prevent mRNA from diffusing back through the central channel into the nucleus<sup>69</sup>, the exact point of first interaction between Dbp5 and mRNA is also unclear<sup>81–84</sup>. Specific binding sites for Dbp5 have been identified in Nup214. Because this is a cytoplasmic filament Nup, it places Dbp5 in an ideal position to receive mRNPs as they begin to exit the NPC, and the remodelling function of Nup214 would thus prevent the mRNPs from re-entering. This model was recently supported by crystal structures of the yeast Dbp5–Gle1–Nup159 (Nup214 in mammals) complex that support Dbp5 binding to RNA. Separation of the carboxy- and amino-terminal RecA-like domains of Dbp5 is triggered by Gle1 in



**Table 1 | The dynamic range of NPC-mediated transport**

Substrate	Dwell time (ms)	Peak centre (nm)	Distribution	Condition	Reference
NTF2	5.8 ± 0.2	−30	Symmetrical	Permeabilized cells	10
NTF2–cargo	5.2 ± 0.2	ND	ND	Permeabilized cells	10
Transportin	7.2 ± 0.3	−2	Symmetrical	Permeabilized cells	10
Transportin–cargo	5.6 ± 0.2	ND	ND	Permeabilized cells	10
Importin-α–cargo (depleted of CAS and GTP)	28 ± 1	ND	ND	Permeabilized cells	50
2×GFP–NLS	10 ± 1	ND	ND	Permeabilized cells, glycerol	39
2×GFP–NLS (depleted of Ran and GTP)	45 ± 5	ND	ND	Permeabilized cells, glycerol	39
2×GFP–NLS (15 mM importin-β)	1.0 ± 0.1	ND	ND	Permeabilized cells, glycerol	44
2×GFP–NLS	7.8 ± 0.4	ND	ND	Living cells, microinjection	44
2×GFP–NLS (competition with dextran)	1.8 ± 0.1	ND	ND	Living cells, microinjection	44
Importin-α–cargo	7.6 ± 0.5	ND	ND	Permeabilized cells	50
Importin-α–cargo (depleted of Ran–GTP)	31 ± 6	ND	ND	Permeabilized cells	50
Ran	10.5 ± 0.8 to 24.8 ± 1.6	−9 ± 82 to −37 ± 82	Symmetrical	Permeabilized cells	36
eGFP	0.4 ± 0.1 to 0.9 ± 0.2	ND	ND	Permeabilized cells	36
BSA	6.2 ± 0.3	−13 ± 1	Symmetrical	Living cells, microinjection	19
Importin-α	7.5 ± 0.8	−6 ± 2	Symmetrical	Living cells, microinjection	19
Importin-β	6.6 ± 0.2*	−10 ± 2	Symmetrical	Living cells, microinjection	19
Importin-β (ΔN44)	11.8 ± 0.6*	−8 ± 1	Symmetrical	Living cells, microinjection	19
Transportin	4.6 ± 0.1*	5 ± 2	Symmetrical	Living cells, microinjection	19
Importin-β	5 ± 2.2	ND	ND	Living cells, microinjection	37
Quantum dots	2 s to 15 min, median 34 s	−5‡	Symmetrical	Permeabilized cells	49
Dys mRNA	500	ND	ND	Living cells, MS2 system	40
β-Actin mRNA	180 ± 10	−97 ± 17 to 71 ± 22	Bimodal	Living cells, MS2 system	11

The dwell times for different factors used to probe NPC transport are given. Errors are indicated as published. Where available, the centre of the binding-site distribution along the transport axis is reported, and the shape of that distribution indicated. The peak centre was measured relative to a POM121–fluorescent-marker fusion protein (either POM121–GFP or POM–tandemTomato). Symmetrical refers to shapes that have one peak and roughly similar decays on both sides. Bimodal refers to β-actin mRNA, for which several binding sites have been found. Condition refers to the preparation of cells and buffer conditions, as discussed in the text. BSA, bovine serum albumin; CAS, recycling cofactor for importin-α; Dys, dystrophin; eGFP, enhanced green fluorescent protein; GFP, green fluorescent protein; ND, not determined; 2×GFP–NLS is an artificial transport cargo molecule, made from a fusion of two GFP molecules that have an NLS. The MS2 system is a method of visualizing mRNA using a cassette of stem–loops that binds tightly to the MS2 coat protein fused to GFP<sup>11</sup>.

\*A second component of ~5 to 15% with a significantly longer dwell time was found.

‡No POM121 used; peak positions found ~20 nm into the central channel.

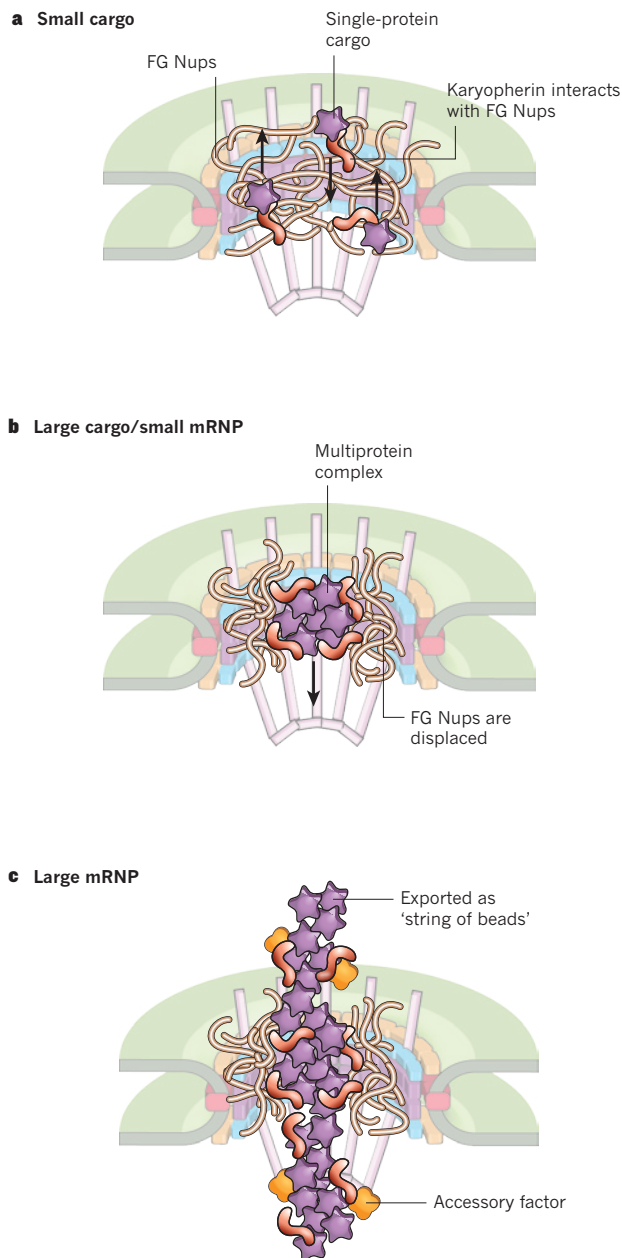
an ATP-dependent manner. After RNA release, Dbp5 is bound by Nup159, resulting in a further separation of the RecA-like domains<sup>64</sup>. Inositol hexakisphosphate binding to Gle1 has been shown to be specific and essential for this process, and a single Dbp5 seems to be able to allow multiple cycles of mRNP remodelling<sup>85,86</sup>. DEAD-box helicases are involved in several nuclear processes that lead to the formation of export-competent mRNPs<sup>68,75,87,88</sup>. Taken together, it seems likely that a certain size limit exists above which rearranging of the mRNP before or during export is mandatory. It also seems safe to speculate that, based on the extensive heterogeneity of mRNAs, this size limit is not sharply defined.

The complete protein content of mRNPs is unknown, so the range of composition differences between different mRNPs is still uncharacterized. Which proteins of the mRNP are involved in mediating transport across the NPC and how many of them are exchanged at the NPC remain central questions in the field. Another key issue is whether a common export mechanism exists for all mRNPs or whether there are transcript-specific differences. In addition, mRNA complexes also have pivotal roles in the life cycle of the cell and are therefore controlled by many processing and checkpoint steps, which are now suspected of being NPC coupled<sup>70</sup>. Molecular crowding<sup>47,89</sup>, discussed before in the context of the molecular environment within the central channel of the NPC (Fig. 2), also has a profound effect on nuclear structure and so could influence the passage of nascent mRNPs to the NPC<sup>48</sup>. For example, it remains unclear whether access to NPCs is sometimes hindered by chromatin, although current super-resolution microscopy data do not suggest this<sup>11,40,90</sup>.

## The dynamics of mRNP export

An insight into the effects that large cargoes may have on transport dynamics is based on imaging quantum dots as they are imported through the NPC from the cytoplasm to the nucleoplasm of living cells<sup>49</sup>. Not surprisingly, transport times were found to be long compared with single protein import measurements. Translocation times of 2 s to several minutes, with a median at 34 s, were measured, which are far longer than those found for the export of similarly sized β-actin mRNPs<sup>11,49</sup> (see also below). This can be explained in part by the fact that quantum dots are rigid substrates and, compared with mRNA complexes, lack the ability to reconfigure during transport. It may also point to the idea that the specific machineries recruited to the mRNP are crucial for ensuring its speedy, as well as specific, transit across the NPC.

Recently, a rather more detailed picture comprising docking, translocation and release for mRNA export across the NPC has been presented<sup>11</sup>. Pivotal for the measurement of nanometre-scale distances between mRNA and NPC was super-registration of the two spectrally resolved signals (Fig. 4). By using the NPCs themselves to generate the registration signal, it was possible to super-register the co-localization of single-molecule signals with ~10 nm precision along the nuclear envelope in the living cell<sup>11</sup>. This detailed picture of mRNA export complements that described previously<sup>40</sup>, in which a model RNA was transiently expressed and its movement traced in the nucleoplasm and during translocation using single-molecule tracking. The translocation time was estimated to be 1 s, based on the data acquisition rate of 1-s time intervals. On the basis of statistical



**Figure 3 | Transport of cargoes.** The challenges faced by the NPC in transporting cargoes of different sizes are shown. Small cargoes are easily accounted for by all existing models (see Fig. 2), but large cargoes raise issues for the functionality of the NPC. **a**, Small cargoes are usually single proteins. They attach to karyopherins, which carry the cargo through the NPC by interacting with the FG Nups. No large-scale displacement of the FG Nups is necessary, and the cargo–karyopherin complexes can be transported bidirectionally. **b**, Large cargoes and RNPs are usually multiprotein complexes that contain several transport factors. Large cargoes displace the FG Nups and sterically hinder other transport. **c**, An mRNP is exported as a 'string of beads', in which each 'bead' behaves as a large cargo. Multiple accessory factors aid in the processing of the mRNP at both the nuclear basket and the cytoplasmic filaments.

analysis of single-molecule tracking data, a diffusion coefficient of  $\sim 0.2\text{--}0.6\ \mu\text{m}^2\ \text{s}^{-1}$  was calculated, and the translocation velocity given as  $0.65\ \mu\text{m}\ \text{s}^{-1}$ . Complex kinetics were inaccessible owing to limitations in the image-acquisition rate, and details of the export step were not observed with this time resolution, but rather acquired through model-based data analysis. Importantly, despite the different sizes of the mRNAs and the very different data-acquisition timescales (3.3-kb  $\beta$ -actin mRNA imaged with 50 Hz, and 4.8-kb mRNA imaged with

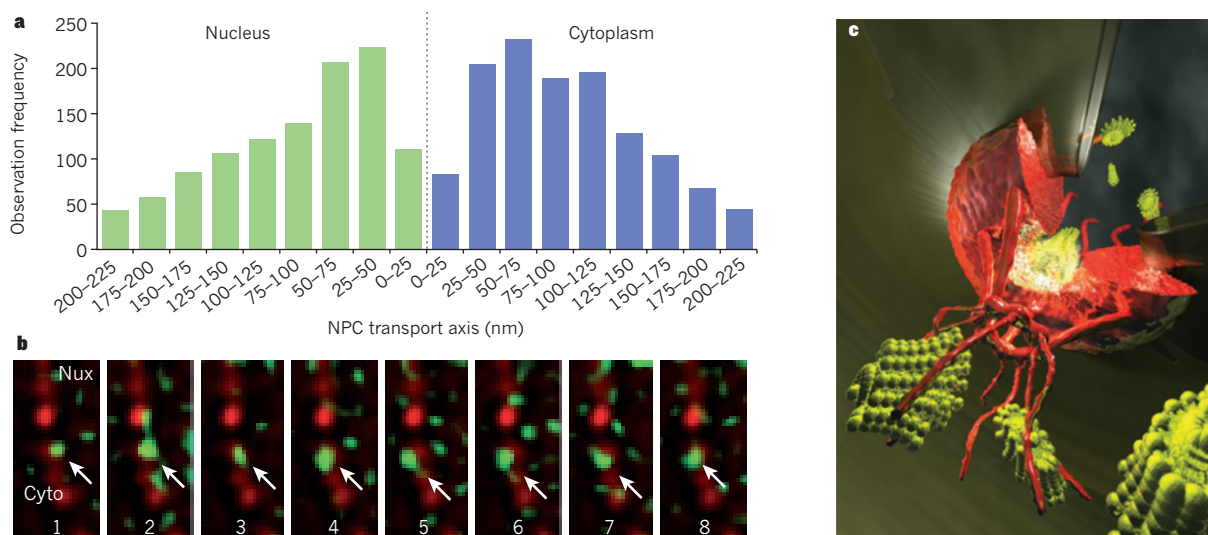
2 Hz), both studies support the rapid transport of mRNAs.

Perhaps the most surprising result to emerge<sup>11</sup> was that a medium-sized endogenous mRNP of about 3 MDa spends most of its transport time of  $\sim 200$  ms equally between docking and release at the nuclear basket and cytoplasmic filaments, whereas translocation through the central channel itself occurs in a remarkably rapid manner within a time interval of less than 20 ms (Fig. 4). This would correspond to a diffusion coefficient along the central channel of roughly  $0.06\ \mu\text{m}^2\ \text{s}^{-1}$  (free diffusion over a 50-nm distance within 10 ms), or a velocity of  $5\ \mu\text{m}\ \text{s}^{-1}$  (linear movement across the central channel). This compares favourably with transport times for protein cargoes that have been found to range from 1 to 15 ms (corresponding to a diffusion coefficient in the channel of  $0.13\ \mu\text{m}^2\ \text{s}^{-1}$ , assuming a 5-ms dwell time that is attributed only to the central channel), and with the free diffusion rate of such a 3 MDa cargo<sup>10,11,39,44,60</sup>. Thus, it seems that the export of RNAs is not limited by getting through the central channel of the NPC, but rather by the time taken in preparation for this transport, and conversely its termination from it. This is analogous to protein import, in which the transport step is minor compared with the assembly of the transport-factor–cargo complex<sup>18</sup>. Given the apparent complexity of the assembling, NPC targeting and disassembling of mRNP cargoes (each consisting of up to hundreds of individual molecules), this makes sense. In the quantum dot study<sup>49</sup>, these docking and release steps were not observed<sup>49</sup>. This could be explained by a much slower translocation step that 'hid' more complex fast kinetics at the rim of the NPC, but also seems to suggest that transport of mRNPs includes steps to hold the mRNP at the docking and release sites. The rapid transition through the central channel must be taken into account when considering which of the transport models is correct. To achieve these times, a model is needed that allows the barrier forces in the central channel and FG Nup region to be overcome within a very short time. It is also clear from these data that mRNP export was not limited in rate by the translocation step, but rather was dependent on the interaction between the cargo and the peripheral elements (at both the nuclear and the cytoplasmic interfaces). This is an important notion as deletion experiments in yeast have shown that most asymmetrical or peripheral Nups are either redundant or unnecessary to achieve selectivity, although the factors associated with some of these proteins (such as Gle1) are important<sup>6,85</sup>. However, it has been shown in yeast that certain types of FG Nup, and not just those associated with the nuclear basket or cytoplasmic filaments, are crucial for efficient mRNP export<sup>6,91</sup>. This indicates that, as with karyopherin-mediated protein transport, particular kinds of FG Nup cooperate to form specific pathways across the NPC that are favoured by specific types of transport-factor–cargo complex<sup>22</sup>.

### Export of ribosomes and other RNAs

Our understanding of the mechanisms and dynamics of the export of other RNAs remains sketchy. Other RNAs include those much smaller than typical mRNAs: for example, transfer RNA, microRNA (miRNA) and small nuclear RNA, but also large RNA-containing particles such as viral RNA, ribosomal RNAs and ribosomal subunits<sup>55,92</sup>. Results indicate that the export of small RNAs is similar to the export of proteins and even involves the same or similar karyopherin transport factors<sup>55</sup>. Both tRNAs and miRNAs seem to carry sequences (or structural elements) analogous to NESs that are recognized by their cognate export karyopherins, whereas mature small nuclear RNP complexes have an NES-containing protein recognized by the export karyopherin CRM1 (ref. 55). Ribosomal subunit export is another topic of great interest<sup>92</sup>. Like mRNP export, the export of both the 40S and the 60S ribosomal subunits must be rapid. Although little is known about the export of the 40S subunit, it has been established that the 60S subunit can use many different pathways for export<sup>91</sup>. This has been interpreted as a mechanism to make this a robust process, less sensitive to the cellular stress response or inhibition. However, the overall regulation, transport mechanisms and detailed dynamics of ribosomal export are much less well understood than for mRNP export. One limit here will be devising





**Figure 4 | Imaging of NPC transport events one molecule at a time.** **a**, By localizing cargo relative to the NPC, spatially resolved binding sites can be recorded along the transport axis of the NPC. The histogram represents data on  $\beta$ -actin mRNA transport<sup>11</sup>. The zero position (dotted line) is determined by localizing the position that POM121 is fused to a fluorescent marker. The two peaks, one on the nuclear surface and one on the cytoplasmic surface of the NPC, are interpreted as docking and release sites. **b**, An image series from a single mRNP export event showing  $\beta$ -actin mRNA (green) traversing the NPC

(red). After docking in the nucleoplasm (Nux) in frame 1, the mRNA (arrows) is repeatedly observed along the NPC until, in frame 8, it reaches the cytoplasm (Cyto). The positions are super-registered to the NPC signal and contribute to the data in **a**. **c**, An artist's impression of a large cargo (green) docking and transiting through the NPC (red). Up to a certain size limit (see text), large cargoes dock to the NPC, transit through the central channel relatively fast, then linger before release. The docking and release steps allow remodelling and/or reorientation of large cargoes. Artwork by Tremani, TU Delft.

a consistent labelling strategy for ribosomal subunits that allows specificity of label targeting to a subclass of 40S or 60S subunits, and stays on the subunits during transport.

## Outlook

Recent work provides insight into how cells transport RNPs across the NPC. Although it seems that the constraints of the narrow channel should make export slow, this is not the case. Instead, the cell has specialized structures on the periphery of the NPC that prepare RNPs for a rapid step through its central channel, and for the repackaging of the RNPs for release into the cytoplasm. The relevance of this mechanism probably extends beyond RNPs, because large complexes of proteins may also need to be temporarily restructured for rapid passage.

Much work over the past few decades has been directed at the structure and composition of the NPCs, but with new microscopic approaches it is now possible to overlay this with a kinetic picture, one that is essential to understand the mechanisms involved in transport. Ensemble measurements have not yet been able to describe sufficiently the individual steps of molecular mobility and interaction, spatial-temporal resolution, kinetic parameters and geographical mapping. The ability to study the dynamics of transport processes opens up key questions, such as the role of the peripheral structures of the NPC in transport, because selectivity seems to be mainly achieved in the central channel. Regulatory functions<sup>93</sup>, links to diseases<sup>94,95</sup> and the ageing of NPCs<sup>96</sup> have been established for the NPC under *in vivo* conditions and are mediated by either specific Nups or transport receptors. However, the spatial overlay of these processes within the NPC remains unclear. A picture of distinct transport pathways for specific cargo along the central channel of the NPC is emerging<sup>37,51</sup>, ultimately leading to the question of whether all NPCs are equal. In single-molecule transport studies only small subsets of NPCs in each experiment show activity<sup>11,19</sup>.  $\beta$ -Actin mRNA was also shown to frequently scan NPCs without engaging in transport. This raises questions addressing NPC activity, such as whether the scanning could be due to certain NPCs (or a specific subpopulation) being inaccessible for mRNA transport, or  $\beta$ -actin mRNA transport specifically. This could result from NPCs being too busy to transport alternative RNA cargoes, or these NPCs could be

resting stages, rendering NPCs temporarily inactive. Alternatively, NPCs might be specialized for particular kinds of transport (see above). Another intriguing possibility is that NPCs could reject the passage of mRNPs during a quality-control surveillance step. In yeast, Nup60 has been implicated in a quality-control step for specific mRNAs localized to the bud tip<sup>97</sup>. It has been suggested that the quality control of complex cargoes — for example, nonsense-mediated decay of premature-termination-codon-containing mRNAs — could occur at the NPC<sup>98</sup>, although it is unclear whether the process is completed at the NPC or whether the NPC simply initiates it.

Study of the NPC has implications for infectious diseases, as it may be possible to inhibit viruses such as HIV by tampering with cellular transport pathways<sup>99</sup>. Moreover, although it is by far the most extensively used, the NPC may not be the only method of crossing the nuclear envelope: some viruses, for example, seem to bypass the NPC entirely and bud directly from the nucleoplasm to the cytoplasm<sup>100</sup>.

A key question in the field is how selectivity in the central channel works and copes with a large variety of cargo sizes, including huge mRNP complexes. The surrounding cellular milieu, and the simultaneous docking to the NPC of multiple transport factors and their large and small cargoes, means that the NPC and its vicinity are very crowded places. Because of this, competition between transport factors, cargoes and non-specific vicinal proteins for space and binding sites must strongly modulate the behaviour of the NPC and RNP export. It will be difficult to completely reproduce all these effects *in vitro*, so the new imaging techniques that have literally shed light on mRNP export will be necessary to understand ultimately how it works. ■

1. Franke, W. W. & Scheer, U. The ultrastructure of the nuclear envelope of amphibian oocytes: a reinvestigation. *J. Ultrastruct. Res.* **30**, 288–316 (1970).
2. Walde, S. & Kehlenbach, R. H. The Part and the Whole: functions of nucleoporins in nucleocytoplasmic transport. *Trends Cell Biol.* **20**, 461–469 (2010).
3. Mattaj, J. W. & Englmeier, L. Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* **67**, 265–306 (1998).
4. Pemberton, L. F. & Paschal, B. M. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6**, 187–198 (2005).
5. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).

**This study describes an approach to combine different experimental data into a common framework with a defined error, underlining the essential features of NPC architecture.**

6. Strawn, L. A., Shen, T. X., Shulga, N., Goldfarb, D. S. & Wente, S. R. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nature Cell Biol.* **6**, 197–206 (2004).
7. Jovanovic-Talman, T. et al. Artificial nanopores that mimic the transport selectivity of the nuclear pore complex. *Nature* **457**, 1023–1027 (2009).
8. Ris, H. & Malecki, M. High-resolution field emission scanning electron microscope imaging of internal cell structures after Epon extraction from sections: a new approach to correlative ultrastructural and immunocytochemical studies. *J. Struct. Biol.* **111**, 148–157 (1993).
9. Kiseleva, E. et al. Yeast nuclear pore complexes have a cytoplasmic ring and internal filaments. *J. Struct. Biol.* **145**, 272–288 (2004).
10. Kubitscheck, U. et al. Nuclear transport of single molecules: dwell times at the nuclear pore complex. *J. Cell Biol.* **168**, 233–243 (2005).
11. Grünwald, D. & Singer, R. *In vivo* imaging of labelled endogenous  $\beta$ -actin mRNA during nucleocytoplasmic transport. *Nature* **467**, 604–607 (2010).  
**This is the first study to follow a single mRNA in detail through the NPC, showing that overall transport times are fast, ~hundreds of milliseconds, and that docking and release are visible kinetic steps.**
12. Gorlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 607–660 (1999).
13. Paine, P. L., Moore, L. C. & Horowitz, S. B. Nuclear envelope permeability. *Nature* **254**, 109–114 (1975).
14. Keminer, O. & Peters, R. Permeability of single nuclear pores. *Biophys. J.* **77**, 217–228 (1999).
15. Mohr, D., Frey, S., Fischer, T., Guttler, T. & Gorlich, D. Characterisation of the passive permeability barrier of nuclear pore complexes. *EMBO J.* **28**, 2541–2553 (2009).
16. Macara, I. G. Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* **65**, 570–594 (2001).
17. Wente, S. R. & Rout, M. P. The nuclear pore complex and nuclear transport. *Cold Spring Harb. Perspect. Biol.* **2**, a000562 (2010).
18. Timney, B. L. et al. Simple kinetic relationships and nonspecific competition govern nuclear import rates *in vivo*. *J. Cell Biol.* **175**, 579–593 (2006).
19. Dange, T., Grünwald, D., Grünwald, A., Peters, R. & Kubitscheck, U. Autonomy and robustness of translocation through the nuclear pore complex: a single-molecule study. *J. Cell Biol.* **183**, 77–86 (2008).
20. Nachury, M. V. & Weis, K. The direction of transport through the nuclear pore can be inverted. *Proc. Natl Acad. Sci. USA* **96**, 9622–9627 (1999).
21. Kopito, R. B. & Elbaum, M. Reversibility in nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA* **104**, 12743–12748 (2007).
22. Terry, L. J. & Wente, S. R. Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryot. Cell* **8**, 1814–1827 (2009).
23. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl Acad. Sci. USA* **100**, 2450–2455 (2003).
24. Lim, R. Y. et al. Nanomechanical basis of selective gating by the nuclear pore complex. *Science* **318**, 640–643 (2007).
25. Frey, S., Richter, R. P. & Gorlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006).
26. Frey, S. & Gorlich, D. A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* **130**, 512–523 (2007).
27. Eisele, N. B., Frey, S., Piehler, J., Gorlich, D. & Richter, R. P. Ultrathin nucleoporin phenylalanine-glycine repeat films and their interaction with nuclear transport receptors. *EMBO Rep.* **11**, 366–372 (2010).
28. Rout, M. P. et al. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
29. Rout, M. P., Aitchison, J. D., Magnasco, M. O. & Chait, B. T. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* **13**, 622–628 (2003).
30. Peters, R. The nanopore connection to cell membrane unitary permeability. *Traffic* **6**, 199–204 (2005).
31. Yamada, J. et al. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics* **9**, 2205–2224 (2010).
32. Lim, R. Y. et al. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA* **103**, 9512–9517 (2006).
33. Zilman, A., Di Talia, S., Chait, B. T., Rout, M. P. & Magnasco, M. O. Efficiency, selectivity, and robustness of nucleocytoplasmic transport. *PLoS Comput. Biol.* **3**, e125 (2007).
34. Zilman, A. et al. Enhancement of transport selectivity through nano-channels by non-specific competition. *PLoS Comput. Biol.* **6**, e1000804 (2010).
35. Huve, J., Wesselmann, R., Kahms, M. & Peters, R. 4Pi microscopy of the nuclear pore complex. *Biophys. J.* **95**, 877–885 (2008).
36. Kahms, M., Lehrich, P., Huve, J., Sanetra, N. & Peters, R. Binding site distribution of nuclear transport receptors and transport complexes in single nuclear pore complexes. *Traffic* **10**, 1228–1242 (2009).
37. Ma, J. & Yang, W. Three-dimensional distribution of transient interactions in the nuclear pore complex obtained from single-molecule snapshots. *Proc. Natl Acad. Sci. USA* **107**, 7305–7310 (2010).  
**In this study, very high spatial resolution is achieved by a combination of confocal excitation with camera detection and modelling of data, supporting the existence of defined cargo transport routes within the NPC.**
38. Kopito, R. B. & Elbaum, M. Nucleocytoplasmic transport: a thermodynamic mechanism. *HFSP J.* **3**, 130–141 (2009).
39. Yang, W., Gelles, J. & Musser, S. M. Imaging of single-molecule translocation through nuclear pore complexes. *Proc. Natl Acad. Sci. USA* **101**, 12887–12892 (2004).
40. Mor, A. et al. Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nature Cell Biol.* **12**, 543–552 (2010).  
**In this paper, various large exogenous mRNP cargos are followed *in vivo*, and their progress from the transcription site to the NPC is shown to be slow (minutes), whereas nuclear transport is more rapid (seconds).**
41. Feldherr, C. M., Kallenbach, E. & Schultz, N. Movement of a karyophilic protein through the nuclear pores of oocytes. *J. Cell Biol.* **99**, 2216–2222 (1984).
42. Dworetzky, S. I. & Feldherr, C. M. Translocation of RNA-coated gold particles through the nuclear pores of oocytes. *J. Cell Biol.* **106**, 575–584 (1988).
43. Richardson, W. D., Mills, A. D., Dilworth, S. M., Laskey, R. A. & Dingwall, C. Nuclear protein migration involves two steps: rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. *Cell* **52**, 655–664 (1988).
44. Yang, W. & Musser, S. M. Nuclear import time and transport efficiency depend on importin  $\beta$  concentration. *J. Cell Biol.* **174**, 951–961 (2006).
45. Ribbeck, K. & Gorlich, D. Kinetic analysis of translocation through nuclear pore complexes. *EMBO J.* **20**, 1320–1330 (2001).
46. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods* **5**, 159–161 (2008).  
**This study introduces a careful calibration of a simple light shield technique for fluorescence imaging, and is the first direct visualization of the high occupancy of NPCs with several individual transport receptors *in vivo*.**
47. Ellis, R. J. Protein folding — inside the cage. *Nature* **442**, 360–362 (2006).
48. Marenduzzo, D., Finan, K. & Cook, P. R. The depletion attraction: an underappreciated force driving cellular organization. *J. Cell Biol.* **175**, 681–686 (2006).
49. Lowe, A. R. et al. Selectivity mechanism of the nuclear pore complex characterized by single cargo tracking. *Nature* **467**, 600–603 (2010).  
**This paper presents the constraints on large cargo transport for artificial, not deformable, cargo, showing the lower time limit for NPC translocation and the upper limit for cargo diameter.**
50. Sun, C., Yang, W., Tu, L. C. & Musser, S. M. Single-molecule measurements of importin  $\alpha$ -cargo complex dissociation at the nuclear pore. *Proc. Natl Acad. Sci. USA* **105**, 8613–8618 (2008).
51. Fiserova, J., Richards, S. A., Wente, S. R. & Goldberg, M. W. Facilitated transport and diffusion take distinct spatial routes through the nuclear pore complex. *J. Cell Sci.* **123**, 2773–2780 (2010).  
**References 37 and 51 use ultrastructural studies and super-fast freezing of samples to capture cargo within the NPC in intact cells, demonstrating that cargo can travel along specific routes in the NPC.**
52. Peters, R. Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. *Traffic* **6**, 421–427 (2005).
53. Dimitrov, D. I., Milchev, A. & Binder, K. Polymer brushes in cylindrical pores: simulation versus scaling theory. *J. Chem. Phys.* **125**, 34905 (2006).
54. Mehlin, H., Daneholt, B. & Skoglund, U. Translocation of a specific premessenger ribonucleoprotein particle through the nuclear-pore studied with electron-microscope tomography. *Cell* **69**, 605–613 (1992).
55. Köhler, A. & Hurt, E. C. Exporting RNA from the nucleus to the cytoplasm. *Nature Rev. Mol. Cell Biol.* **8**, 761–773 (2007).
56. Akey, C. W. Visualization of transport-related configurations of the nuclear pore transporter. *Biophys. J.* **58**, 341–355 (1990).
57. Iborra, F. J., Jackson, D. A. & Cook, P. R. The path of RNA through nuclear pores: apparent entry from the sides into specialized pores. *J. Cell Sci.* **113**, 291–302 (2000).
58. Siebrasse, J. P. & Kubitscheck, U. Single molecule tracking for studying nucleocytoplasmic transport and intranuclear dynamics. *Methods Mol. Biol.* **464**, 343–361 (2009).
59. Galy, V. et al. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116**, 63–73 (2004).
60. Siebrasse, J. P. et al. Discontinuous movement of mRNP particles in nucleoplasmic regions devoid of chromatin. *Proc. Natl Acad. Sci. USA* **105**, 20291–20296 (2008).  
**This careful analysis of RNP mobility within the nucleus demonstrates that different mobility distributions observed for an RNP are best explained by single molecules alternating between tethering and diffusion.**
61. Kiseleva, E., Goldberg, M. W., Allen, T. D. & Akey, C. W. Active nuclear pore complexes in *Chironomus*: visualization of transporter configurations related to mRNP export. *J. Cell Sci.* **111**, 223–236 (1998).
62. Soop, T. et al. Nup153 affects entry of messenger and ribosomal ribonucleoproteins into the nuclear basket during export. *Mol. Biol. Cell* **16**, 5610–5620 (2005).
63. Dargemont, C. & Kuhn, L. C. Export of mRNA from microinjected nuclei of *Xenopus laevis* oocytes. *J. Cell Biol.* **118**, 1–9 (1992).
64. Montpetit, B. et al. A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP<sub>6</sub> in mRNA export. *Nature* **472**, 238–242 (2011).  
**This study presents the atomic structures of protein complexes for mRNA and factors that have been implicated in NPC-related export, and provides a model for how the release step of large cargo from the NPC is achieved.**
65. Conti, E. & Izaurralde, E. Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.* **13**, 310–319 (2001).



66. Reed, R. & Hurt, E. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108**, 523–531 (2002).
67. Kota, K. P., Wagner, S. R., Huerta, E., Underwood, J. M. & Nickerson, J. A. Binding of ATP to UAP56 is necessary for mRNA export. *J. Cell Sci.* **121**, 1526–1537 (2008).
68. Carmody, S. R. & Wenthe, S. R. mRNA nuclear export at a glance. *J. Cell Sci.* **122**, 1933–1937 (2009).
69. Stewart, M. Ratcheting mRNA out of the nucleus. *Mol. Cell* **25**, 327–330 (2007).
70. Rodriguez-Navarro, S. & Hurt, E. Linking gene regulation to mRNA production and export. *Curr. Opin. Cell Biol.* **23**, 302–309 (2011).
71. Braun, I. C., Herold, A., Rode, M. & Izaurralde, E. Nuclear export of mRNA by TAP/NXF1 requires two nucleoporin-binding sites but not p15. *Mol. Cell Biol.* **22**, 5405–5418 (2002).
72. Segref, A. *et al.* Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)<sup>+</sup> RNA and nuclear pores. *EMBO J.* **16**, 3256–3271 (1997).
73. Li, Y. *et al.* An intron with a constitutive transport element is retained in a *Tap* messenger RNA. *Nature* **443**, 234–237 (2006).
74. Hutten, S. & Kehlenbach, R. H. CRM1-mediated nuclear export: to the pore and beyond. *Trends Cell Biol.* **17**, 193–201 (2007).
75. Schmitt, C. *et al.* Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *EMBO J.* **18**, 4332–4347 (1999).
76. Forler, D. *et al.* RanBP2/Nup358 provides a major binding site for NXF1-p15 dimers at the nuclear pore complex and functions in nuclear mRNA export. *Mol. Cell Biol.* **24**, 1155–1167 (2004).
77. Weirich, C. S. *et al.* Activation of the DEXD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP<sub>6</sub> is required for mRNA export. *Nature Cell Biol.* **8**, 668–676 (2006).
78. Hodge, C. A., Colot, H. V., Stafford, P. & Cole, C. N. Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1-1 cells. *EMBO J.* **18**, 5778–5788 (1999).
79. Lund, M. K. & Guthrie, C. The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Mol. Cell* **20**, 645–651 (2005).
80. Linder, P. mRNA export: RNP remodeling by DEAD-box proteins. *Curr. Biol.* **18**, R297–R299 (2008).
81. Zhao, J., Jin, S. B., Bjorkroth, B., Wieslander, L. & Daneholt, B. The mRNA export factor Dbp5 is associated with Balbiani ring mRNP from gene to cytoplasm. *EMBO J.* **21**, 1177–1187 (2002).
82. Cole, C. N. & Scarcelli, J. J. Transport of messenger RNA from the nucleus to the cytoplasm. *Curr. Opin. Cell Biol.* **18**, 299–306 (2006).
83. Bolger, T. A., Folkmann, A. W., Tran, E. J. & Wenthe, S. R. The mRNA export factor Gle1 and inositol hexakisphosphate regulate distinct stages of translation. *Cell* **134**, 624–633 (2008).
84. von Moeller, H., Basquin, C. & Conti, E. The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nature Struct. Mol. Biol.* **16**, 247–254 (2009).
85. Alcazar-Roman, A. R., Bolger, T. A. & Wenthe, S. R. Control of mRNA export and translation termination by inositol hexakisphosphate requires specific interaction with Gle1. *J. Biol. Chem.* **285**, 16683–16692 (2010).
86. Noble, K. N., Tran, E. J., Alcázar-Román, A. R., Hodge, C. A., Cole, C. N. & Wenthe, S. R. The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes Dev.* **25**, 1065–1077 (2011).
87. Gatfield, D. *et al.* The DEXH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila*. *Curr. Biol.* **11**, 1716–1721 (2001).
88. Stutz, F. & Izaurralde, E. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol.* **13**, 319–327 (2003).
89. Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **11**, 114–119 (2001).
90. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332–1336 (2008).  
**Using fixed cells, this work gives a first glance at the possible contributions of super-resolution microscopy, providing high-resolution images of nuclear structure and showing how NPCs may be made accessible for large cargo.**
91. Terry, L. J. & Wenthe, S. R. Nuclear mRNA export requires specific FG nucleoporins for translocation through the nuclear pore complex. *J. Cell Biol.* **178**, 1121–1132 (2007).
92. Lo, K. Y. & Johnson, A. W. Reengineering ribosome export. *Mol. Biol. Cell* **20**, 1545–1554 (2009).
93. Shitashige, M. *et al.* Regulation of Wnt signaling by the nuclear pore complex. *Gastroenterology* **134**, 1961–1971 (2008).
94. Alvisi, G., Rawlinson, S. M., Ghildyal, R., Ripalti, A. & Jans, D. A. Regulated nucleocytoplasmic trafficking of viral gene products: a therapeutic target? *Biochim. Biophys. Acta* **1784**, 213–227 (2008).
95. Hurt, J. A. & Silver, P. A. mRNA nuclear export and human disease. *Dis. Model Mech.* **1**, 103–108 (2008).
96. D'Angelo, M. A., Raices, M., Panowski, S. H. & Hetzer, M. W. Age-dependent deterioration of nuclear pore complexes causes a loss of nuclear integrity in postmitotic cells. *Cell* **136**, 284–295 (2009).
97. Powrie, E. A., Zenklusen, D. & Singer, R. H. A nucleoporin, Nup60p, affects the nuclear and cytoplasmic localization of *ASH1* mRNA in *S. cerevisiae*. *RNA* **17**, 134–144 (2010).
98. Isken, O. & Maquat, L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* **21**, 1833–1856 (2007).
99. Satterly, N. *et al.* Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **104**, 1853–1858 (2007).
100. Lee, C. P. & Chen, M. R. Escape of herpesviruses from the nucleus. *Rev. Med. Virol.* **20**, 214–230 (2010).

**Acknowledgements** We apologize to those colleagues whose work, through space considerations, could not be discussed or cited in this review. This work has been supported by funds from the Kavli Foundation to D.G., National Institutes of Health grants GM86217 and GM84364 to R.H.S., and GM062427, RR022220 and GM071329 to M.R. We thank A. Joseph for critically reading the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to R.H.S. ([robert.singer@einstein.yu.edu](mailto:robert.singer@einstein.yu.edu)).

# Type VI secretion delivers bacteriolytic effectors to target cells

Alistair B. Russell<sup>1</sup>, Rachel D. Hood<sup>1</sup>, Nhat Khai Bui<sup>2</sup>, Michele LeRoux<sup>3</sup>, Waldemar Vollmer<sup>2</sup> & Joseph D. Mougous<sup>1</sup>

**Peptidoglycan is the major structural constituent of the bacterial cell wall, forming a meshwork outside the cytoplasmic membrane that maintains cell shape and prevents lysis. In Gram-negative bacteria, peptidoglycan is located in the periplasm, where it is protected from exogenous lytic enzymes by the outer membrane. Here we show that the type VI secretion system of *Pseudomonas aeruginosa* breaches this barrier to deliver two effector proteins, Tse1 and Tse3, to the periplasm of recipient cells. In this compartment, the effectors hydrolyse peptidoglycan, thereby providing a fitness advantage for *P. aeruginosa* cells in competition with other bacteria. To protect itself from lysis by Tse1 and Tse3, *P. aeruginosa* uses specific periplasmically localized immunity proteins. The requirement for these immunity proteins depends on intercellular self-intoxication through an active type VI secretion system, indicating a mechanism for export whereby effectors do not access donor cell periplasm in transit.**

Competition for niches among bacteria is widespread, fierce and deliberate. These organisms produce factors ranging in complexity from small diffusible molecules to multicomponent machines, in order to inhibit the proliferation of rival cells<sup>1,2</sup>. A common target of such factors is the peptidoglycan cell wall<sup>3–6</sup>. The conserved, essential and accessible nature of this molecule makes it an Achilles' heel of bacteria.

The type VI secretion system (T6SS) is a complex and widely distributed protein export machine capable of cell-contact-dependent targeting of effector proteins between Gram-negative bacterial cells<sup>7–10</sup>. However, the mechanism by which effectors are delivered via the secretory apparatus, and the function(s) of the effectors within recipient cells, have remained elusive. Current models of the T6SS derive from the observation that several of its components share structural homology to bacteriophage proteins<sup>11–13</sup>; it has been proposed that target cell recognition and effector delivery occur in a process analogous to bacteriophage entry<sup>14</sup>.

The observation that T6S can target bacteria was originally made through studies of the haemolysin co-regulated protein secretion island I (HSI-I)-encoded T6SS (H1-T6SS) of *P. aeruginosa*, which exports at least three proteins: Tse1, Tse2 and Tse3 (refs 7, 13). These proteins are unrelated to each other and lack significant primary sequence homology to characterized proteins. One substrate, Tse2, is toxic by an unknown mechanism in the cytoplasm of recipient cells lacking Tsi2, a Tse2-specific immunity protein.

Here we show that Tse1 and Tse3 are lytic enzymes that degrade peptidoglycan via amidase and muramidase activity, respectively. Unlike related enzymes associated with other secretion systems<sup>15</sup>, these proteins are not required for the assembly of a functional secretory apparatus. Instead, Tse1 and Tse3 function as lytic antibacterial effectors that depend upon T6S to breach the barrier imposed by the Gram-negative outer membrane. Contacting *P. aeruginosa* cells actively intoxicate each other with Tse1 and Tse3. However, the peptidoglycan of *P. aeruginosa* is not inherently resistant to the activities of these enzymes. To protect itself, the bacterium synthesizes immunity proteins—T6S immunity 1 and 3 (Tsi1 and Tsi3)—that specifically interact with and inactivate

cognate toxins in the periplasm. Orthologues of *tsi1* and *tsi3* seem to be restricted to *P. aeruginosa*; therefore, the species is able to exploit the H1-T6SS to target closely related organisms that are likely to compete for overlapping niches, while minimizing the fitness cost associated with self-targeting.

## Tse1 and Tse3 are lytic enzymes

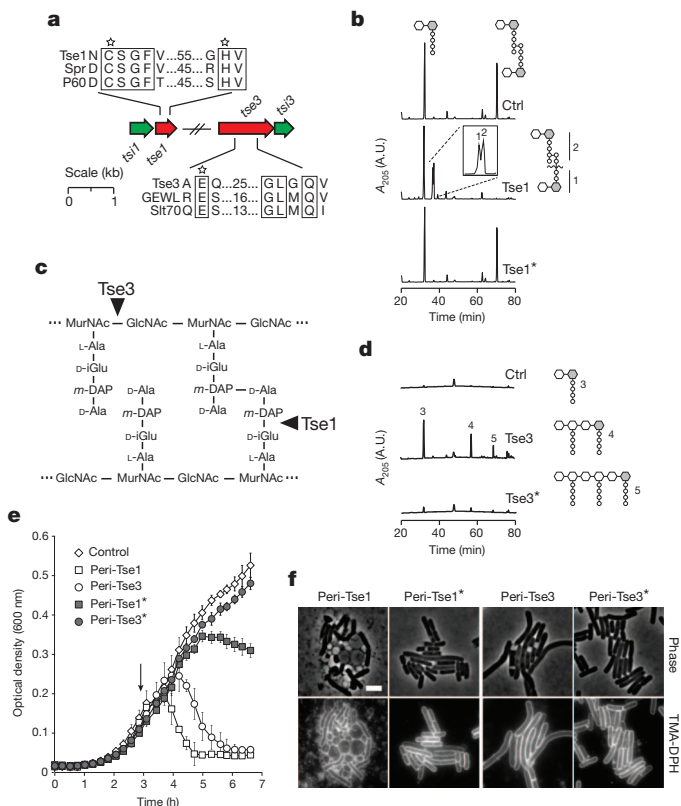
To identify potential functions of Tse1 and Tse3, we searched their sequences for catalytic motifs using structure prediction algorithms<sup>16</sup>. Interestingly, motifs present in peptidoglycan-degrading enzymes were apparent in both proteins (Fig. 1a and Supplementary Fig. 1). Tse1 contains invariant catalytic amino acids present in certain cell wall amidases (DL-endopeptidases)<sup>17</sup>, whereas Tse3 possesses a motif that includes a catalytic glutamic acid found in muramidases<sup>18,19</sup>.

To test our structure-based predictions, we incubated purified Tse1 and Tse3 (Supplementary Fig. 2) with isolated *Escherichia coli* peptidoglycan sacculi. Soluble products released by the enzymes were separated by high-performance liquid chromatography (HPLC) and analysed by mass spectrometry (MS). To generate separable fragments, Tse1-treated samples were digested with cellosyl, a muramidase, before HPLC. The observed absence of the major crosslinked fragment, and the formation of two Tse1-specific products, is consistent with enzymatic cleavage of an amide bond in the peptidoglycan peptide crosslink (Fig. 1b and Supplementary Fig. 3). Moreover, our MS data indicate that the enzyme possesses specificity for the  $\gamma$ -D-glutamyl-L-meso-diaminopimelic acid bond in the donor peptide stem (Fig. 1c and Supplementary Table 1). A variant of Tse1 containing an alanine substitution in its predicted catalytic cysteine (C30A, hereafter called Tse1\*) did not degrade peptidoglycan (Fig. 1b).

Soluble peptidoglycan fragments released by Tse3 confirmed our prediction that the enzyme cleaves the glycan backbone between N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) residues (Fig. 1d and Supplementary Fig. 3). Enzymes that cleave this bond can do so hydrolytically (lysozymes) or non-hydrolytically (lytic transglycosylases); the latter results in the formation of 1,6-anhydroMurNAc. Our analyses showed that Tse3 possesses lysozyme-like activity and suggest that its activity is limited to a

<sup>1</sup>Department of Microbiology, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. <sup>3</sup>Molecular and Cellular Biology Program, University of Washington, Seattle, Washington 98195, USA.





**Figure 1 | Tse1 and Tse3 are lytic proteins belonging to amidase and muramidase enzyme families.** **a**, Genomic organization of *tse1* and *tse3* and their homology with characterized amidase and muramidase enzymes, respectively. Highly conserved (boxed) and catalytic (starred) residues of the respective enzyme families are indicated. SWISS-PROT entry names for the proteins shown are: GEWL, LYG\_ANSAN; P60, P60\_LISIN; Slt70, SLT\_ECOLI; Spr, SPR\_ECOLI; Tse1, Q9I2Q1\_PSEAE; Tse3, Q9HYC5\_PSEAE. See Supplementary Fig. 1 for full alignments. **b**, **d**, Partial HPLC chromatograms of sodium borohydride-reduced soluble *E. coli* peptidoglycan products resulting from digestion with Tse1 and subsequent cleavage with cellosyl (b) or digestion with Tse3 alone (d). Peak assignments were made based on MS; predicted structures are shown schematically with hexagons and circles corresponding to sugars and amino acid residues, respectively. Reduced sugar moieties are shown with grey fill. Full chromatograms and MS data are provided in Supplementary Fig. 3 and Supplementary Table 1.  $A_{205}$ , absorbance at 205 nm; A.U., arbitrary units. **c**, Simplified representation of Gram-negative peptidoglycan showing cleavage sites of Tse1 and Tse3 based on data summarized in b and d. **e**, Growth in liquid media of *E. coli* producing the indicated peri-Tse proteins. Periplasmic localization was achieved by fusion to the PelB leader sequence<sup>35</sup>. Cultures were induced at the indicated time (arrow). Error bars indicate  $\pm$ s.d. ( $n = 3$ ). **f**, Representative micrographs of strains shown in e acquired before complete lysis. The lipophilic dye TMA-DPH is used to highlight the cellular membranes. Supplementary Fig. 5 contains the full microscopic fields from which these images were derived. All images were acquired at the same magnification. Scale bar: 2  $\mu$ m.

fraction of the MurNAc–GlcNAc bonds. The enzyme solubilized a significant proportion of the sacculi to release non-crosslinked peptidoglycan fragments and high-molecular-weight, soluble peptidoglycan fragments (Fig. 1c, Supplementary Fig. 3 and Supplementary Table 1). A Tse3 protein with glutamine substituted at the site of the predicted catalytic glutamic acid (E250Q, hereafter called Tse3\*) displayed significantly diminished activity.

If Tse1 and Tse3 degrade peptidoglycan, we reasoned that the enzymes might have the capacity to lyse bacterial cells. Ectopic expression of Tse1 and Tse3 in the cytoplasm of *E. coli* resulted in no significant lysis (Supplementary Fig. 4a, b). However, periplasmically localized forms of both proteins (peri-Tse1, peri-Tse3) abruptly lysed cells after induction (Fig. 1e and Supplementary Fig. 4c). In accordance

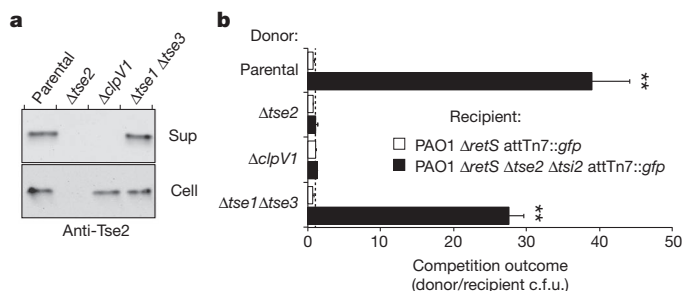
with our *in vitro* studies, peri-Tse1\* and peri-Tse3\* did not induce lysis at expression levels equivalent to those of the native enzymes (Supplementary Fig. 4d). We also examined cells producing the periplasmically localized enzymes using fluorescence microscopy. Consistent with our biochemical data, cells producing peri-Tse1 were amorphous or spherical, whereas those producing peri-Tse3 were swollen and filamentous (Fig. 1f and Supplementary Fig. 5). In total, these data demonstrate that Tse1 and Tse3 are enzymes that degrade peptidoglycan *in vivo*, and that, unlike related enzymes involved in cell wall metabolism, they possess no inherent means of accessing their substrate in the periplasmic space.

## T6S function does not require Tse1 and Tse3

Because the Tse enzymes alone are unable to reach their target cellular compartment, we hypothesized that their function must be linked to export by the T6SS. In this regard, they could: 1) remodel donor peptidoglycan to allow for the assembly of the mature T6S apparatus; 2) remodel recipient cell peptidoglycan to facilitate the passage of the T6S apparatus through the recipient cell wall; or 3) act as antibacterial effectors that compromise recipient cell wall integrity. To determine if Tse1 and Tse3 are essential for T6S apparatus assembly, we examined whether the enzymes are required for export of the third effector, Tse2. The secretion of Tse2 was not diminished in a strain lacking *tse1* and *tse3*, indicating that assembly of the T6S apparatus is unhindered by their absence (Fig. 2a). If Tse1 and Tse3 act as enzymes that remodel recipient cell peptidoglycan to facilitate effector translocation, Tse2 action on recipient cells should be severely impaired or nullified in the  $\Delta tse1 \Delta tse3$  background. Instead, we found that this strain retained the ability to functionally target Tse2 to recipient cells (Fig. 2b). These findings led us to examine further the hypothesis that Tse1 and Tse3 are effector proteins rather than accessory enzymes of the T6S apparatus.

## Immunity proteins inhibit Tse1 and Tse3

Previous data indicate that *P. aeruginosa* can target itself via the T6SS (ref. 7). If Tse1 and Tse3 act as antibacterial effectors, it follows that *P. aeruginosa* must be immune to their toxic effects. The *tse1* and *tse3* genes are each found in predicted bicistronic operons with a hypothetical gene, henceforth referred to as *tsi1* and *tsi3*, respectively (Fig. 1a). Immunity proteins often inactivate their cognate toxin by direct interaction<sup>20</sup>; therefore, as a first step towards defining a functional link between cognate Tsi and Tse proteins, we asked whether they physically associate. A solution containing a mixture of purified Tse1 and Tse3 was mixed with *E. coli* lysates containing either Tsi1 or

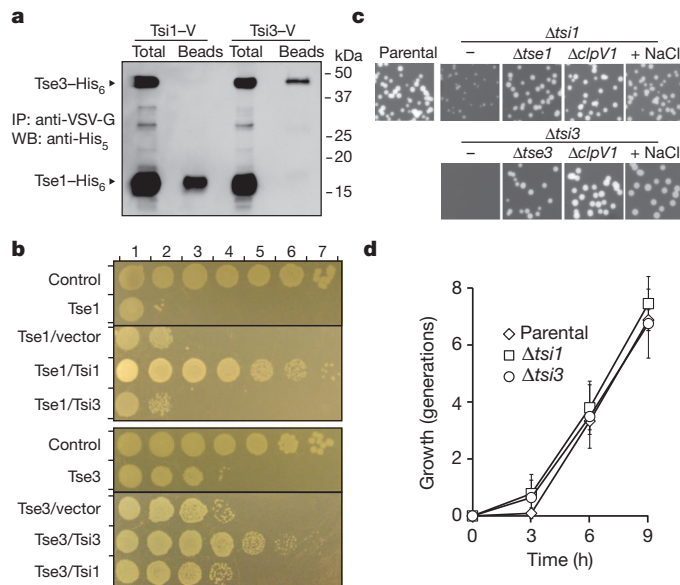


**Figure 2 | Tse1 and Tse3 are not required for Tse2 export or transfer to recipient cells via the T6S apparatus.** **a**, Western blot analysis of supernatant (Sup) and cell-associated (Cell) fractions of the indicated *P. aeruginosa* strains. The parental background for all experiments represented in this figure is PAO1  $\Delta retS$ , a strain in which the H1-T6SS is activated constitutively<sup>13,36</sup>. **b**, Growth competition assays between the indicated donor and recipient strains under T6S-conductive conditions. Experiments were initiated with equal colony forming units (c.f.u.) of donor and recipient bacteria as denoted by the dashed line. The  $\Delta clpV1$  strain is a T6S-deficient control. Asterisks indicate significant differences in competition outcome between recipient strains against the same donor strain. \*\* $P < 0.01$ . Error bars indicate  $\pm$ s.d. ( $n = 3$ ).

**Tsi3.** Co-immunoprecipitation studies indicated that Tsi1 and Tsi3 interact specifically with Tse1 and Tse3, respectively, and interactions between non-cognate pairs were not detected (Fig. 3a). To investigate the immunity properties of the Tsi proteins, we measured their ability to inhibit toxicity of peri-Tse1 and peri-Tse3 in *E. coli*. Both Tsi1 and Tsi3 significantly decreased the toxicity of cognate, but not non-cognate, Tse proteins (Fig. 3b). These results show that the activity of periplasmic Tse1 and Tse3 is specifically inhibited by cognate Tsi proteins.

### T6S delivers Tse1 and Tse3 to the periplasm

Most genes encoding immunity functions are essential in the presence of their cognate toxins. However, mutations that inactivate *tsi1* and *tsi3* are readily generated in *P. aeruginosa* strains that constitutively express and export Tse1 and Tse3. On the basis of this observation, we hypothesized that under standard laboratory conditions, the Tse proteins do not efficiently access their substrate in the periplasm. This suggests that T6S occurs by a mechanism wherein effectors are denied access to donor cell periplasm and are instead released directly to the periplasm of the recipient cell. According to this mechanism, the *tsi* genes would only be essential when a strain is grown under conditions that permit intercellular transfer of effectors between neighbouring cells by the T6SS. As predicted, deletions in *tsi1* and *tsi3* severely impaired the growth of *P. aeruginosa* on a solid substrate, a condition conducive to T6S-based effector delivery (Fig. 3c and Supplementary Fig. 6)<sup>21,22</sup>. In contrast, this growth inhibition did not occur in liquid media, which is not conducive to effector delivery by the T6SS (Fig. 3d). The growth inhibition phenotype required a functional T6SS and intact cognate effector genes, and consistent with the proposed functions of Tse1 and Tse3 in compromising cell wall integrity, growth of immunity-deficient strains was fully rescued by increasing the osmolarity of the medium (Fig. 3c).



**Figure 3 | Tsi1 and Tsi3 provide immunity to cognate toxins.** **a**, Western blot analysis of hexahistidine-tagged Tse proteins (–His<sub>6</sub>) in total and bead-associated fractions of an anti-VSV-G (vesicular stomatitis virus glycoprotein) immunoprecipitation of VSV-G epitope fused Tsi proteins (–V) from *E. coli*. **b**, Growth of *E. coli* harbouring a vector expressing the indicated *tse* gene (top panels) or vectors expressing the indicated *tse* and *tsi* genes (bottom panels). Numbers at the top indicate tenfold serial dilutions. **c**, Fluorescence micrographs showing colony growth of the indicated strains. The parental background for this experiment was PAO1  $\Delta retS$  attTn7::gfp. Growth of the  $\Delta tsi$  strains was rescued by the addition of 1.0% w/v NaCl to the underlying medium. For quantification of data and complementation analyses see Supplementary Fig. 7. **d**, Replication rates of the indicated *P. aeruginosa* strains in liquid medium of low osmolarity formulated as in c. The parental strain used in this experiment was PAO1  $\Delta retS$ . Error bars indicate  $\pm$  s.d. ( $n = 3$ ).

Bioinformatic analyses suggested that the Tsi proteins reside in the periplasm—Tsi1 as a soluble periplasmic protein and Tsi3 as an outer membrane lipoprotein. These predictions were confirmed by subcellular fractionation experiments, which indicated enrichment of the proteins in the periplasmic compartment (Fig. 4a). This result, taken together with the observation that the Tsi proteins interact directly with their cognate Tse proteins (Fig. 3a), provided us with a means of addressing whether the T6SS delivers Tse proteins intercellularly to the periplasm. We reasoned that if the Tse proteins are indeed delivered to the periplasm of another bacterial cell, not only should we be able to observe intoxication between distinct donor and recipient strains of *P. aeruginosa*, but the production of an otherwise competent immunity protein that is mislocalized to the cytoplasm should not be able to prevent such intoxication.

In growth competition assays between distinct donor and recipient strains of *P. aeruginosa*, we found that recipient cells that lack Tse3 immunity and are incapable of self-intoxication ( $\Delta tse3 \Delta tsi3$ ) display a growth disadvantage against donor bacteria (Fig. 4b). This phenotype depends on H1-T6SS function and Tse3 in the donor strain. In the recipient strain, ectopic expression of wild-type *tsi3*, but not an allele encoding a signal-sequence-deficient protein (Tsi3-SS), rescues the fitness defect. Importantly, the Tsi3-SS protein used in this experiment does not reach the periplasm, and retains activity *in vitro* as judged by interaction with Tse3 (Fig. 4a and Supplementary Fig. 7). The Tsi3-SS protein also fails to rescue the intercellular self-intoxication growth phenotype of  $\Delta tsi3$  (Supplementary Fig. 6). Analogous experiments with Tsi1 were not feasible, as the protein was unstable in the cytoplasm.

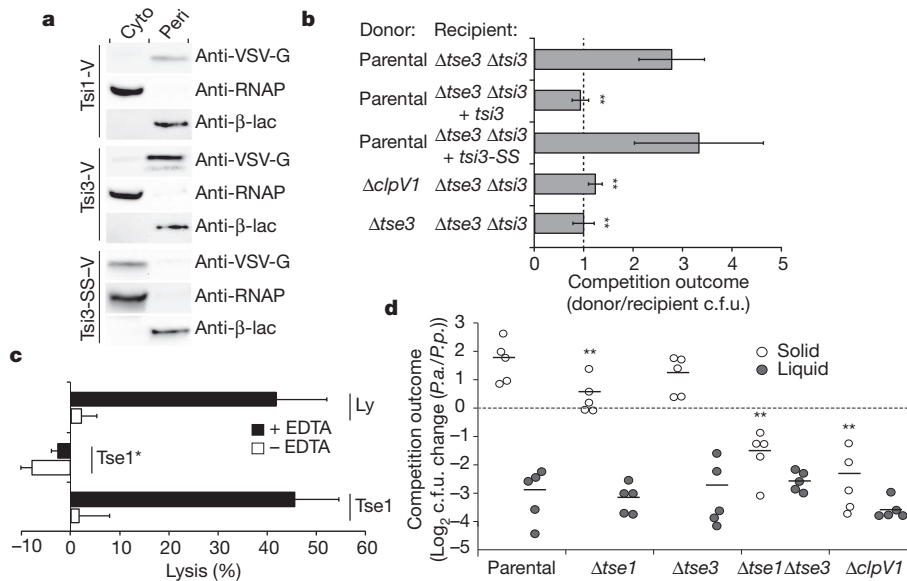
The most parsimonious explanation for T6S-mediated intercellular toxicity by Tse1 and Tse3 is that the apparatus provides a conduit for the effectors through the outer membrane of recipient cells. This led us to predict that exogenous Tse1 and Tse3 would not lyse intact *P. aeruginosa*. Furthermore, we posited that if the outer membrane was the relevant barrier to Tse1 and Tse3 toxicity, compromising its integrity should render *P. aeruginosa* susceptible to exogenous administration of the enzymes.

To test these predictions, we measured lysis of permeabilized and intact *P. aeruginosa* after addition of exogenous Tse1. We did not test Tse3, as the filamentous phenotype induced by this enzyme would not affect non-growing, permeabilized cells. Intact *P. aeruginosa* cells were not affected by the addition of exogenous Tse1; conversely, permeabilized *P. aeruginosa* was highly susceptible to lysis by the enzyme (Fig. 4c). Lysis induced by Tse1 is linked to its enzymatic function, as Tse1\* failed to lyse cells significantly. In total, our data show that the T6SS breaches the outer membrane to deliver lytic effector proteins directly to recipient cell periplasm.

### The H1-T6SS targets effectors to *P. putida*

To determine whether the T6SS can target the Tse proteins to cells of another Gram-negative organism, we conducted growth competition assays between *P. aeruginosa* and *Pseudomonas putida*. These bacteria can be co-isolated from the environment<sup>23</sup> and are likely to compete for niches<sup>24</sup>. Whereas inactivation of either *tse1* or *tse3* only modestly affected the outcome of *P. aeruginosa*–*P. putida* competition assays, the fitness of *P. aeruginosa* lacking both genes or a functional T6SS was markedly impaired (Fig. 4a). This partial redundancy is congruent with the enzymes exerting their effects through a single target (peptidoglycan) in the recipient cell. The fitness advantage provided by Tse1 and Tse3 was lost in liquid medium, consistent with cell-contact-dependent delivery of the proteins to competitor cells (Fig. 4d). These data indicate that the T6SS targets its effectors to other species of bacteria and that these proteins can be key determinants in the outcome of interspecies bacterial interactions. In contrast with intraspecies intoxication, interspecies intoxication via the T6SS does not require the inactivation of a negative regulator of the system (for example,  $\Delta retS$ ), indicating that T6S function is stimulated in response to rival bacteria.





**Figure 4 | Tse1 and Tse3 delivered to the periplasm provide a fitness advantage to donor cells.** **a**, Western blot analyses of cytoplasmic (Cyto) and periplasmic (Peri) fractions of *P. aeruginosa* strains producing Tsi1-V, Tsi3-V or Tsi3-SS-V. Equivalent ratios of the cyto and peri samples were loaded in each panel. RNA polymerase (RNAP) and  $\beta$ -lactamase ( $\beta$ -lac) enzymes were used as cytoplasmic and periplasmic fractionation controls, respectively. The presence of Tsi3, a predicted outer membrane lipoprotein, in the periplasmic fraction is consistent with previous studies using this method of fractionation<sup>37</sup>. **b**, Growth competition assays between the indicated donor and recipient strains under T6S-conductive conditions. Experiments were initiated with equal c.f.u. of donor and recipient bacteria as denoted by the dashed line. The parental strain used in this experiment was PAO1  $\Delta retS$ . All donor strains were modified at the attB site with *lacZ*. Asterisks indicate outcomes significantly different than parental versus

## Discussion

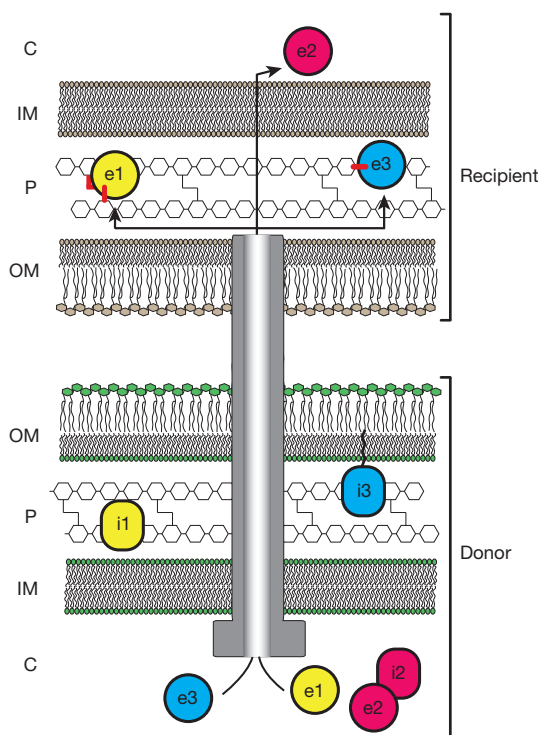
Our data lead us to propose a model for T6S-catalysed translocation of effectors to the periplasm of recipient bacteria (Fig. 5). This model provides a mechanistic framework for understanding the form and

function of this complex secretion system. Our findings strengthen the existing hypothesis that the T6SS is evolutionarily and functionally related to bacteriophage<sup>8,14,25</sup>. Neither the T4 bacteriophage tail spike nor other components of the puncturing device are thought to cross the inner membrane; instead, bacteriophage DNA is released to the periplasm and subsequently enters the cytoplasmic compartment using another pathway<sup>26</sup>. By analogy, the Tse proteins would use T6S components as a puncturing device to gain access to the periplasm, whereupon Tse2 may then utilize an independent route to access the cytoplasm (Fig. 5).

Niche competition in natural environments has clearly selected for potent antibacterial processes; however, the human body is also home to a complex and competitive microbiota<sup>27,28</sup>. Commensal bacteria form a protective barrier, and the ability of pathogens to colonize the host is not only dependent on suppression or subversion of host immunity, but can also depend on their ability to displace these more innocuous organisms<sup>29–31</sup>. In polymicrobial infections, Gram-negative bacteria, including *P. aeruginosa*, often compete with other Gram-negative bacteria for access to nutrient-rich host tissue<sup>32</sup>. Factors such as the T6SS, that influence the relative fitness of these organisms, are thus likely to have an impact on disease outcome.

Our data lead us to propose a model for T6S-catalysed translocation of effectors to the periplasm of recipient bacteria (Fig. 5). This model provides a mechanistic framework for understanding the form and function of this complex secretion system. Our findings strengthen the existing hypothesis that the T6SS is evolutionarily and functionally related to bacteriophage<sup>8,14,25</sup>. Neither the T4 bacteriophage tail spike nor other components of the puncturing device are thought to cross the inner membrane; instead, bacteriophage DNA is released to the periplasm and subsequently enters the cytoplasmic compartment using another pathway<sup>26</sup>. By analogy, the Tse proteins would use T6S components as a puncturing device to gain access to the periplasm, whereupon Tse2 may then utilize an independent route to access the cytoplasm (Fig. 5).

Niche competition in natural environments has clearly selected for potent antibacterial processes; however, the human body is also home to a complex and competitive microbiota<sup>27,28</sup>. Commensal bacteria form a protective barrier, and the ability of pathogens to colonize the host is not only dependent on suppression or subversion of host immunity, but can also depend on their ability to displace these more innocuous organisms<sup>29–31</sup>. In polymicrobial infections, Gram-negative bacteria, including *P. aeruginosa*, often compete with other Gram-negative bacteria for access to nutrient-rich host tissue<sup>32</sup>. Factors such as the T6SS, that influence the relative fitness of these organisms, are thus likely to have an impact on disease outcome.



**Figure 5 | Proposed mechanism of T6S-dependent delivery of effector proteins.** The schematic depicts the junction between competing bacteria, with a donor cell delivering the Tse effector proteins through the T6S apparatus (grey tube) to recipient cell periplasm. Effector and immunity proteins are shown as circles and rounded rectangles, respectively. Bonds in the peptidoglycan that are predicted targets of the effector proteins are highlighted (red). The cytoplasm (C), inner membrane (IM), periplasm (P) and outer membrane (OM) of both bacteria are shown.

## METHODS SUMMARY

*P. aeruginosa* strains used in this study were derived from the sequenced strain PAO1 (ref. 33). All deletions were in-frame and unmarked, and were generated by allelic exchange. *E. coli* growth curves were conducted using BL21 pLysS cells harbouring expression plasmids for *tse* and *tsi* genes. Intercellular self-intoxication and interbacterial competition assays were performed by spotting mixed overnight cultures on a nitrocellulose membrane placed on a 3% agar growth medium. Samples were incubated at 37 °C (*P. aeruginosa*–*P. aeruginosa*) or 30 °C (*P. aeruginosa*–*P. putida*) for 12 or 24 h. Tse1-catalysed *P. aeruginosa* lysis was measured by placing cells in a minimal buffer  $\pm$  1.5 mM EDTA containing either Tse1, Tse1\* or lysozyme. The change in optical density at 600 nm after 5 min of incubation was used to calculate lysis. For determination of Tse1 and Tse3 activity, isolated *E. coli* peptidoglycan sacculi were incubated with the purified enzymes (100  $\mu$ g ml<sup>-1</sup>). The resulting peptidoglycan and soluble fragments released by the enzymes were separated by HPLC and their identities were determined using MS as described previously<sup>34</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 18 February; accepted 23 May 2011.**

- Hayes, C. S., Aoki, S. K. & Low, D. A. Bacterial contact-dependent delivery systems. *Annu. Rev. Genet.* **44**, 71–90 (2010).
- Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Rev. Microbiol.* **8**, 15–25 (2010).
- Gründling, A. & Schneewind, O. Cross-linked peptidoglycan mediates lysostaphin binding to the cell wall envelope of *Staphylococcus aureus*. *J. Bacteriol.* **188**, 2463–2472 (2006).
- Vollmer, W., Pils, H., Hantke, K., Höltje, J. V. & Braun, V. Pesticin displays muramidase activity. *J. Bacteriol.* **179**, 1580–1583 (1997).
- Brötz, H., Bierbaum, G., Markus, A., Molitor, E. & Sahl, H. G. Mode of action of the lantibiotic mersacidin: inhibition of peptidoglycan biosynthesis via a novel mechanism? *Antimicrob. Agents Chemother.* **39**, 714–719 (1995).
- Riley, M. A. & Wertz, J. E. Bacteriocins: evolution, ecology, and application. *Annu. Rev. Microbiol.* **56**, 117–137 (2002).
- Hood, R. D. et al. A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**, 25–37 (2010).
- Schwarz, S., Hood, R. D. & Mougous, J. D. What is type VI secretion doing in all those bugs? *Trends Microbiol.* **18**, 531–537 (2010).
- Cascales, E. The type VI secretion toolkit. *EMBO Rep.* **9**, 735–741 (2008).
- Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide *in silico* analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**, 104 (2009).
- Ballister, E. R., Lai, A. H., Zuckermann, R. N., Cheng, Y. & Mougous, J. D. *In vitro* self-assembly of tailorable nanotubes from a simple protein building block. *Proc. Natl Acad. Sci. USA* **105**, 3733–3738 (2008).
- Leiman, P. G. et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
- Mougous, J. D. et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526–1530 (2006).
- Kanamaru, S. Structural similarity of tailed phages and pathogenic bacterial secretion systems. *Proc. Natl Acad. Sci. USA* **106**, 4067–4068 (2009).
- Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. & Cascales, E. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu. Rev. Microbiol.* **59**, 451–485 (2005).
- Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
- Anantharaman, V. & Aravind, L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol.* **4**, R11 (2003).
- Scheurwater, E., Reid, C. W. & Clarke, A. J. Lytic transglycosylases: bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **40**, 586–591 (2008).
- Vollmer, W., Joris, B., Charlier, P. & Foster, S. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol. Rev.* **32**, 259–286 (2008).
- Gerdes, K., Christensen, S. K. & Lobner-Olesen, A. Prokaryotic toxin-antitoxin stress response loci. *Nature Rev. Microbiol.* **3**, 371–382 (2005).
- Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. *Nature Rev. Microbiol.* **2**, 95–108 (2004).
- Schwarz, S. et al. *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog.* **6**, (2010).
- Mortensen, J. E., Fisher, M. C. & LiPuma, J. J. Recovery of *Pseudomonas cepacia* and other *Pseudomonas* species from the environment. *Infect. Control Hosp. Epidemiol.* **16**, 30–32 (1995).
- Nelson, K. E. et al. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799–808 (2002).
- Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl Acad. Sci. USA* **104**, 15508–15513 (2007).
- Rakhuba, D. V., Kolomiets, E. I., Dey, E. S. & Novik, G. I. Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Pol. J. Microbiol.* **59**, 145–155 (2010).
- Nelson, K. E. et al. A catalog of reference genomes from the human microbiome. *Science* **328**, 994–999 (2010).
- Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- Brook, I. Bacterial interference. *Crit. Rev. Microbiol.* **25**, 155–172 (1999).
- Iwase, T. et al. *Staphylococcus epidermidis* Esp inhibits *Staphylococcus aureus* biofilm formation and nasal colonization. *Nature* **465**, 346–349 (2010).
- Reid, G., Howard, J. & Gan, B. S. Can bacterial interference prevent infection? *Trends Microbiol.* **9**, 424–428 (2001).
- Gjedsbøl, K. et al. Multiple bacterial species reside in chronic wounds: a longitudinal study. *Int. Wound J.* **3**, 225–231 (2006).
- Stover, C. K. et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964 (2000).
- Bui, N. K. et al. The peptidoglycan sacculus of *Myxococcus xanthus* has unusual structural features and is degraded during glycerol-induced myxospore development. *J. Bacteriol.* **191**, 494–505 (2009).
- Lei, S. P., Lin, H. C., Wang, S. S., Callaway, J. & Wilcox, G. Characterization of the *Erwinia carotovora pelB* gene and its product pectate lyase. *J. Bacteriol.* **169**, 4379–4383 (1987).
- Goodman, A. L. et al. A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*. *Dev. Cell* **7**, 745–754 (2004).
- Imperi, F. et al. Analysis of the periplasmic proteome of *Pseudomonas aeruginosa*, a metabolically versatile opportunistic pathogen. *Proteomics* **9**, 1901–1915 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Singh, E. Nester, H. Kulasekara, N. Salama, E. P. Greenberg, L. Ramakrishnan and members of the Mougous laboratory for discussions and critical reading of the manuscript, the Harwood laboratory for use of their microscope, and J. Gray of the Pinnacle Laboratory of Newcastle University for MS analysis. This work was supported by the National Institutes of Health (J.D.M.; RO1 AI080609) and the European Commission within the DIVINOCELL programme (W.V.). A.B.R. was supported by a Graduate Research Fellowship from the National Science Foundation.

**Author Contributions** A.B.R., R.D.H., N.K.B., M.L., W.V. and J.D.M. conceived and designed experiments. A.B.R., R.D.H., N.K.B. and J.D.M. conducted experiments. A.B.R., R.D.H., W.V. and J.D.M. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.D.M. (mougous@u.washington.edu).



## METHODS

**Bacterial strains, plasmids and growth conditions.** *P. aeruginosa* strains used in this study were derived from the sequenced strain PAO1 (ref. 33). *P. aeruginosa* strains were grown on either Luria-Bertani media (LB), or the equivalent lacking additional NaCl (LB low salt (LB-LS)): 10 g bactopeptone and 5 g yeast extract per litre) at 37 °C supplemented with 30 µg ml<sup>-1</sup> gentamycin, 25 µg ml<sup>-1</sup> irgasan, 5% w/v sucrose, 40 µg ml<sup>-1</sup> X-gal, and stated concentrations of IPTG as required. *E. coli* strains in this study included DH5α for plasmid maintenance, SM10 for conjugal transfer of plasmids into *P. aeruginosa*, BL21 pLysS for expression of Tse1 and Tse3 for toxicity and lysis, and Shuffle T7 pLysS Express (New England Biolabs), for purification of Tse1 and Tse3. All *E. coli* strains were grown on either LB or LB-LS at 37 °C supplemented with 15 µg ml<sup>-1</sup> gentamycin, 150 µg ml<sup>-1</sup> carbenicillin, 50 µg ml<sup>-1</sup> kanamycin, 30 µg ml<sup>-1</sup> chloramphenicol, 200 µg ml<sup>-1</sup> trimethoprim, 0.1% rhamnose, and stated concentrations of IPTG as required. *P. putida* used in this study was the sequenced strain, KT2440 (ref. 24). *P. putida* was grown on LB or LB-LS at 30 °C. In all experiments where expression from a plasmid was required, strains were grown on media supplemented with required antibiotics to select for plasmid maintenance.

Plasmids used for inducible expression were pPSV35CV for *P. aeruginosa* and pET29b+ (Novagen), pET22b+ (Novagen), pSCRhaB2<sup>38</sup> and pPSV35CV for *E. coli*<sup>39</sup>. Chromosomal deletions were made as described previously<sup>40</sup>.

**DNA manipulations.** The creation, maintenance and transformation of plasmid constructs followed standard molecular cloning procedures. All primers used in this study were obtained from Integrated DNA Technologies. DNA amplification was carried out using either Phusion (New England Biolabs) or Mangomix (Bioline). DNA sequencing was performed by Genewiz Incorporated. Restriction enzymes were obtained from New England Biolabs. SOE PCR was performed as previously described<sup>41</sup>.

**Plasmid construction.** pPSV35CV, pEXG2 and pSCRhaB2 have been described previously<sup>38–40</sup>. *E. coli* pET29+ expression vectors for Tse1 and Tse3 were constructed by standard cloning techniques following amplification from PAO1 chromosomal DNA using the primer pairs 1289/1290 and 1291/1292, respectively. *E. coli* pET22b+ expression vectors for Tse1 and Tse3 were constructed in a similar manner using primer pairs 1477/1478 and 1475/1476. Point mutations were introduced using Quikchange (Stratagene) with primer pairs 1479/1480 and 1481/1482 for the production of *tse1*(C30A) and *tse3*(E250Q), respectively.

pPSV35CV expression vectors for Tsi1 and Tsi3 were generated by amplifying the genes from genomic DNA using primer pairs 1469/1470 and 1472/1473, respectively. The Tsi3-SS pPSV35CV expression vector was generated from a product amplified using the primer pair 1522/1473. The pSCRhaB2 vectors for expressing Tsi proteins in *E. coli* were produced by amplifying the genes using primer pairs 1470/1497 for *tse1* and 1473/1498 for *tse3*. A VSV-epitope tag was then cloned downstream of these two genes for the purpose of tagged expression.

All deletions were in-frame and were generated by exchange with deletion alleles constructed by SOE PCR. For *tse1*, *tse3*, *tse1* and *tse3* deletion constructs, upstream DNA flanking sequences were amplified by 628/629, 735/736, 721/722 and 1485/1486, respectively. Downstream flanking DNA sequences were amplified by 630/631, 737/738, 723/724 and 1487/1488, respectively. Deletions of both effector and immunity protein were accomplished by amplifying upstream regions of *tse1-tsi1* and *tse3-tsi3* with 721/722 and 735/736 respectively and downstream regions with 628/629 and 835/836 respectively.

**Growth curves.** For *E. coli* growth curves, BL21 pLysS cells harbouring expression plasmids were grown overnight in liquid LB shaking at 37 °C and subinoculated to a starting optical density at 600 nm (OD<sub>600</sub>) of between 0.01 and 0.02 in LB-LS. Cultures were grown to OD<sub>600</sub> 0.1–0.2 and induced with 0.1 mM IPTG. The vector pET29b+ was used for expression of native Tse1 and Tse3, and the pET22b+ vector was used for expression of periplasmic Tse1 and Tse3, and catalytic amino acid substitutions thereof. Both vectors added a carboxy-terminal hexahistidine tag to expressed proteins, allowing for western blot analysis of expression. Samples for western blot analysis were taken 30 min after induction for Tse1, peri-Tse1 and peri-Tse1\* and 45 min after induction for Tse3, peri-Tse3 and peri-Tse3\*.

For *P. aeruginosa* growth curves, cells were grown overnight at 37 °C in liquid LB with shaking and sub-inoculated 1:1,000 into LB-LS. Growth was measured by enumerating c.f.u. from plate counts of samples taken at the indicated time points.

***E. coli* toxicity measurements.** Overnight LB cultures of *E. coli* harbouring pET22b+ expression vectors and *E. coli* harbouring both pET22b+ and pSCRhaB2 expression vectors were serially diluted in LB to 10<sup>6</sup> as tenfold dilutions. These dilutions were spotted onto LB-LS agar with the following concentrations of inducer molecules: 0.075 mM IPTG for pET22b+::tse1, pET22b+::tse3 and the associated vector control, 0.02 mM IPTG and 0.1% rhamnose for pET22b+::tse1 pSCRhaB2::tse1 and all associated controls, and 0.05 mM IPTG and 0.1% rhamnose

for pET22b+::tse3 pSCRhaB2::tse3 and all associated controls. Pictures were taken between 20 and 26 h after plating.

**Subcellular fractionation.** *P. aeruginosa* Δ*retS* cells harbouring expression vectors for Tsi1–V, Tsi3–V, or Tsi3-SS–V and an additional vector expressing TEM-1 (pPSV18) were grown overnight. This overnight culture was sub-inoculated into LB supplemented with 0.1 mM IPTG and grown to late logarithmic phase. Periplasmic and cytoplasmic fractions were prepared as described<sup>137,42</sup>.

*E. coli* BL21 cells harbouring expression vectors for Tse1\*, Tse3\*, peri-Tse1\* and peri-Tse3\* were grown overnight and sub-inoculated into LB. For Tse1\* and Tse3\*, fractionation cells also carried an empty pET22b vector to provide expression of TEM-1. Cells were grown to an OD<sub>600</sub> of 0.1 and induced with either 0.1 mM IPTG (Tse1\* and peri-Tse1\*) or 0.5 mM IPTG (Tse3\* and peri-Tse3\*). Cells were then harvested and fractionated as described<sup>43</sup>.

**Preparation of proteins and western blotting.** Cell-associated and supernatant samples were prepared as described previously<sup>39</sup>. Western blotting was performed as described previously for anti-VSV-G and anti-RNA polymerase<sup>13</sup> with the modification that anti-VSV-G antibody probing was performed in 5% BSA in Tris-buffered saline containing 0.05% v/v Tween 20. The anti-Tse2 polyclonal rabbit antibody was raised against the peptide YDGDVGRYLHPDKEC (GenScript). Western blots using both this antibody and the α-β-lactamase antibody (QED Biosciences Inc.) were performed identically to those using anti-VSV-G. The anti-His<sub>5</sub> western blots were performed using the Penta-His HRP Conjugate Kit according to manufacturer's instructions (Qiagen).

**Immunoprecipitation.** BL21 pLysS cells expressing VSV-G-tagged Tsi1, Tsi3, or Tsi3-SS were pelleted and re-suspended in lysis buffer (20 mM Tris-Cl pH 7.5, 50 mM KCl, 8.0% v/v glycerol, 0.1% v/v NP 40, 1.0% v/v triton, supplemented with Dnase I (Roche), lysozyme (Roche), and Sigmafast protease inhibitor (Sigma) according to the manufacturer's instructions). Cells were disrupted by sonication to release VSV-G-tagged Tsi proteins into solution. To this suspension, Tse1 and Tse3 were added to concentrations of 30 µg ml<sup>-1</sup> and 25 µg ml<sup>-1</sup>, respectively. This mixture was clarified by centrifugation, and a sample of the supernatant was taken as a pre-immunoprecipitation sample. The remainder of the supernatant was incubated with 100 µl anti-VSV-G agarose beads (Sigma) for 2 h at 4 °C. Beads were washed three times with IP-wash buffer (100 mM NaCl, 25 mM KCl, 0.1% v/v triton, 0.1% v/v NP-40, 20 mM Tris-Cl pH 7.5, and 2% v/v glycerol). Proteins were removed from beads with SDS loading buffer (125 mM Tris, pH 6.8, 2% (w/v) 2-mercaptoethanol, 20% (v/v) glycerol, 0.001% (w/v) bromophenol blue and 4% (w/v) SDS) and analysed by western blot.

**Interbacterial competition assays.** The inter-*P. aeruginosa* competitions were performed as described previously with minor modifications<sup>7</sup>. For experiments described in both Fig. 2b and Fig. 4b, competition assays were performed on nitrocellulose on LB or LB-LS 3% agar, respectively. Plate counts were taken of the initial inoculum to ensure a starting c.f.u. ratio of 1:1, and again after either 24 h (Fig. 2b) or 12 h (Fig. 4b) to obtain a final c.f.u. ratio. Donor and recipient colonies were disambiguated through fluorescence imaging (Fig. 2e) or through the activity of a β-galactosidase reporter as visualized on plates containing 40 µg ml<sup>-1</sup> X-gal (Fig. 4b)<sup>5</sup>. Data were analysed using a two-tailed Student's *t*-test.

For interspecies competition assays, cultures of *P. aeruginosa* and *P. putida* were grown overnight in LB broth at 37 °C and 30 °C, respectively. Cultures were then washed in LB and re-suspended to an OD<sub>600</sub> of 4.0 for *P. aeruginosa* and 4.5 for *P. putida*. *P. putida* and *P. aeruginosa* were mixed in a one-to-one ratio by volume, this mixture was spotted on a nitrocellulose membrane placed on LB-LS 3% agar, and the c.f.u. ratio of the organisms was measured by plate counts. The assays were incubated for 24 h at 30 °C, after which the cells were re-suspended in LB broth and the final c.f.u. ratio determined through plate counts. Data were analysed using a one-tailed Student's *t*-test.

**Purification of Tse1 and Tse3.** For purification, Tse1, Tse3, Tse1\* and Tse3\* were expressed in pET29b+ vectors in Shuffle Express T7 lysY cells (New England Biolabs). The proteins were purified to homogeneity using previously reported methods<sup>44</sup>, except that in all steps no reducing agents or lysozyme were used.

**Bioinformatic analyses.** Predicted structural homology was queried using PHYRE<sup>16</sup>. Alignments were performed using T-Espresso<sup>45</sup>. Sequences of cell wall amidases and muramidases for alignments were obtained from seed sequences from PFAM<sup>46</sup>. Critical motifs were defined by previous work in the study of NlpC/P60 and lytic transglycosylase/GEWL enzymes<sup>17,18</sup>.

**Enzymatic assays.** Tse1 and Tse1\*: purified peptidoglycan sacculi (300 µg) from *E. coli* MC1061 (ref. 47) were incubated with Tse1 or Tse1\* (100 µg ml<sup>-1</sup>) in 300 µl of 20 mM Tris/HCl, pH 8.0 for 4 h at 37 °C. A sample with enzyme buffer instead of Tse1 served as a control. The pH was adjusted to 4.8 and the sample was incubated with 40 µg ml<sup>-1</sup> of the muramidase cellosyl (provided by Höchst AG) for 16 h at 37 °C to convert the residual peptidoglycan or solubilized fragments into muropeptides. The sample was boiled for 10 min and insoluble material was removed by brief centrifugation. The reduced muropeptides were reduced with

sodium borohydride and analysed by HPLC as described<sup>47</sup>. Fractions 1 and 2 were collected, concentrated in a SpeedVac, acidified by 1% trifluoroacetic acid and analysed by offline electrospray mass spectrometry on a Finnigan LTQ-FT mass spectrometer (ThermoElectron) as described<sup>34</sup>.

Tse3 and Tse3\*: purified peptidoglycan sacculi (300 µg) from *E. coli* MC1061 were incubated with Tse3 or Tse3\* (100 µg ml<sup>-1</sup>) in 300 µl of 20 mM sodium phosphate, pH 4.8 for 20 h at 37 °C. A sample with enzyme buffer instead of Tse3 served as a control. The samples were boiled for 10 min and centrifuged for 15 min (16,000g). The supernatant was reduced with sodium borohydride and analysed by HPLC as described above (supernatant samples). The pellet was re-suspended in 20 mM sodium phosphate, pH 4.8 and incubated with 40 µg ml<sup>-1</sup> cellosyl for 14 h at 37 °C. The samples were boiled for 10 min, cleared by brief centrifugation and analysed by HPLC as described above (pellet samples). Fractions 3, 4 and 5 were collected and analysed by mass spectrometry as described above.

**Self-intoxication assays.** PAO1  $\Delta retS$  attTn7::gfp cells bearing the indicated gene deletions were grown overnight in LB broth at 37 °C. Cells were then diluted to 10<sup>3</sup> c.f.u. ml<sup>-1</sup> and 20 µl of this solution was placed on a nitrocellulose membrane placed on LB-LS 3% agar or LB 3% agar (contains 1.0% w/v NaCl). Fluorescence images were acquired following 23 h of incubation at 37 °C. For quantification and complementation, non-fluorescent strains were used and 1 mM IPTG was included for induction of all strains—except for the *tsi3*-complemented strain, for which no IPTG was required to achieve comparable levels of expression to the *tsi3*-SS-complemented strain. At 23 h cells were re-suspended in LB. Plate counts of the initial inoculum and the final suspension were used to determine growth. Data were analysed using a one-tailed Student's *t*-test.

**Fluorescence microscopy.** BL21 pLysS cells harbouring periplasmic expression vectors for Tse1, Tse3 and catalytic substitution mutants were grown in conditions identical to those in the *E. coli* growth curve experiments. Cells were harvested 30 min after induction for Tse1 experiments and 1 h after induction for Tse3 experiments. These cells were re-suspended in PBS and incubated with 0.3 µM TMA-DPH (1-(4-trimethylammoniumphenyl)-6-phenyl-1,3,5-hexatriene *p*-toluenesulphonate) for 10 min. The stained cells were placed on 1% agarose pads containing PBS for microscopic analysis. Microscopy was performed as described previously<sup>39</sup>.

**EDTA-permeabilization lysis assay.** Assays were performed as previously described with minor modifications<sup>48</sup>. Cells were sub-inoculated into LB broth from overnight liquid cultures and grown to late logarithmic phase. Cells were washed in 20 mM Tris-Cl pH 7.5 and Tse1, Tse1\*, or lysozyme were added to a final concentration of 0.01 mg ml<sup>-1</sup>. An initial OD<sub>600</sub> measurement was taken before EDTA pH 8.0 was added to a final concentration of 1.5 mM. Cells were incubated with shaking at 37 °C for 5 min and a final OD<sub>600</sub> reading was taken. *P. aeruginosa* undergoes rapid autolysis under these assay conditions, thus lysis was expressed as a percentage of lysis above a buffer-only control.

38. Cardona, S. T. & Valvano, M. A. An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid* **54**, 219–228 (2005).
39. Hsu, F., Schwarz, S. & Mougous, J. D. TagR promotes PpkA-catalysed type VI secretion activation in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **72**, 1111–1125 (2009).
40. Rietsch, A., Vallet-Gely, I., Dove, S. L. & Mekalanos, J. J. ExsE, a secreted regulator of type III secretion genes in *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA* **102**, 8006–8011 (2005).
41. Horton, R. M. *et al.* Gene splicing by overlap extension. *Methods Enzymol.* **217**, 270–279 (1993).
42. Wood, P. M. Periplasmic location of the terminal reductase in nitrite respiration. *FEBS Lett.* **92**, 214–218 (1978).
43. Liu, J. & Walsh, C. T. Peptidyl-prolyl cis-trans-isomerase from *Escherichia coli*: a periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. *Proc. Natl Acad. Sci. USA* **87**, 4028–4032 (1990).
44. Mougous, J. D. *et al.* Identification, function and structure of the mycobacterial sulfotransferase that initiates sulfolipid-1 biosynthesis. *Nature Struct. Mol. Biol.* **11**, 721–729 (2004).
45. Armougom, F. *et al.* Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34**, W604–W608 (2006).
46. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (2010).
47. Glauner, B. Separation and quantification of mucopeptides with high-performance liquid chromatography. *Anal. Biochem.* **172**, 451–464 (1988).
48. Watt, S. R. & Clarke, A. J. Role of autolysins in the EDTA-induced lysis of *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* **124**, 113–119 (1994).



# An integrated semiconductor device enabling non-optical genome sequencing

Jonathan M. Rothberg<sup>1</sup>, Wolfgang Hinz<sup>1</sup>, Todd M. Rearick<sup>1</sup>, Jonathan Schultz<sup>1</sup>, William Mileski<sup>1</sup>, Mel Davey<sup>1</sup>, John H. Leamon<sup>1</sup>, Kim Johnson<sup>1</sup>, Mark J. Milgrew<sup>1</sup>, Matthew Edwards<sup>1</sup>, Jeremy Hoon<sup>1</sup>, Jan F. Simons<sup>1</sup>, David Marran<sup>1</sup>, Jason W. Myers<sup>1</sup>, John F. Davidson<sup>1</sup>, Annika Branting<sup>1</sup>, John R. Nobile<sup>1</sup>, Bernard P. Puc<sup>1</sup>, David Light<sup>1</sup>, Travis A. Clark<sup>1</sup>, Martin Huber<sup>1</sup>, Jeffrey T. Branciforte<sup>1</sup>, Isaac B. Stoner<sup>1</sup>, Simon E. Cawley<sup>1</sup>, Michael Lyons<sup>1</sup>, Yutao Fu<sup>1</sup>, Nils Homer<sup>1</sup>, Marina Sedova<sup>1</sup>, Xin Miao<sup>1</sup>, Brian Reed<sup>1</sup>, Jeffrey Sabina<sup>1</sup>, Erika Feierstein<sup>1</sup>, Michelle Schorn<sup>1</sup>, Mohammad Alanjary<sup>1</sup>, Eileen Dimalanta<sup>1</sup>, Devin Dressman<sup>1</sup>, Rachel Kasinskas<sup>1</sup>, Tanya Sokolsky<sup>1</sup>, Jacqueline A. Fidanza<sup>1</sup>, Eugeni Namsaraev<sup>1</sup>, Kevin J. McKernan<sup>1</sup>, Alan Williams<sup>1</sup>, G. Thomas Roth<sup>1</sup> & James Bustillo<sup>1</sup>

**The seminal importance of DNA sequencing to the life sciences, biotechnology and medicine has driven the search for more scalable and lower-cost solutions. Here we describe a DNA sequencing technology in which scalable, low-cost semiconductor manufacturing techniques are used to make an integrated circuit able to directly perform non-optical DNA sequencing of genomes. Sequence data are obtained by directly sensing the ions produced by template-directed DNA polymerase synthesis using all-natural nucleotides on this massively parallel semiconductor-sensing device or ion chip. The ion chip contains ion-sensitive, field-effect transistor-based sensors in perfect register with 1.2 million wells, which provide confinement and allow parallel, simultaneous detection of independent sequencing reactions. Use of the most widely used technology for constructing integrated circuits, the complementary metal-oxide semiconductor (CMOS) process, allows for low-cost, large-scale production and scaling of the device to higher densities and larger array sizes. We show the performance of the system by sequencing three bacterial genomes, its robustness and scalability by producing ion chips with up to 10 times as many sensors and sequencing a human genome.**

DNA sequencing and, more recently, massively parallel DNA sequencing<sup>1–4</sup> has had a profound impact on research and medicine. The reductions in cost and time for generating DNA sequence have resulted in a range of new sequencing applications in cancer<sup>5,6</sup>, human genetics<sup>7</sup>, infectious diseases<sup>8</sup> and the study of personal genomes<sup>9–11</sup>, as well as in fields as diverse as ecology<sup>12,13</sup> and the study of ancient DNA<sup>14,15</sup>. Although *de novo* sequencing costs have dropped substantially, there is a desire to continue to drop the cost of sequencing at an exponential rate consistent with the semiconductor industry's Moore's Law<sup>16</sup> as well as to provide lower cost, faster and more portable devices. This has been operationalized by the desire to reach the \$1,000 genome<sup>17</sup>.

To date, DNA sequencing has been limited by its requirement for imaging technology, electromagnetic intermediates (either X-rays<sup>18</sup>, or light<sup>19</sup>) and specialized nucleotides or other reagents<sup>20</sup>. To overcome these limitations and further democratize the practice of sequencing, a paradigm shift based on non-optical sequencing on newly developed integrated circuits was pursued. Owing to its scalability and its low power requirement, CMOS processes are dominant in modern integrated circuit manufacturing<sup>21</sup>. The ubiquitous nature of computers, digital cameras and mobile phones has been made possible by the low-cost production of integrated circuits in CMOS.

Leveraging advances in the imaging field—which has produced large, fast arrays for photonic imaging<sup>22</sup>—we sought a suitable electronic sensor for the construction of an integrated circuit to detect the hydrogen ions that would be released by DNA polymerase<sup>23</sup> during sequencing by synthesis, as opposed to a sensor designed for the detection of photons. Although a variety of electrochemical detection methods have been studied<sup>24,25</sup>, the ion-sensitive field-effect transistor (ISFET)<sup>26,27</sup> was most applicable to our chemistry and scaling requirements because of

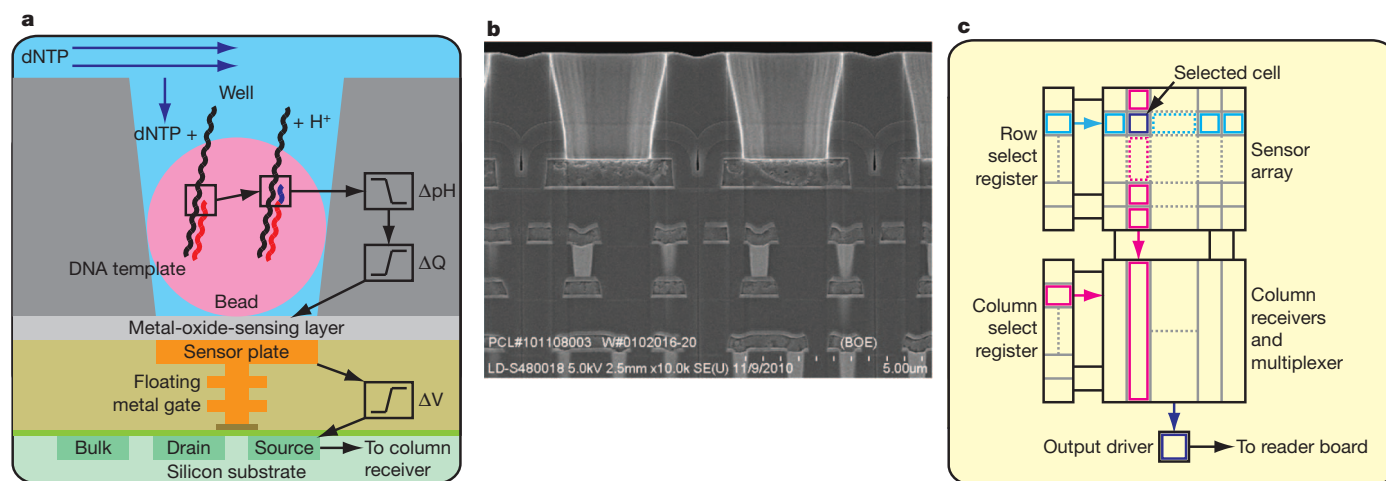
its sensitivity to hydrogen ions, and its compatibility with CMOS processes<sup>28–31</sup>. Previous attempts to detect both single-nucleotide polymorphisms (SNPs)<sup>32</sup> and DNA synthesis<sup>33</sup> as well as sequence DNA electronically<sup>34</sup> have been made. However, none of them produced *de novo* DNA sequence, addressed the issue of delivering template DNA to the sensors, or scaled to large arrays. In addition, previous efforts in ISFETs were limited in the number of sensors per array, the yield of working independent sensors and readout speed<sup>35,36</sup>, and encountered difficulty in exposing the sensors to fluids while protecting the electronics<sup>37</sup>.

Here, we overcome previous limitations with electronic detection and enable the production of chips with a large number of fast, uniform, working sensors. Our focus has been on the development of these ion chips, as well as the biochemical methods, supporting instrumentation and software needed to enable *de novo* DNA sequencing for applications requiring millions to billions of bases (Supplementary Fig. 1). A typical 2-h run using an ion chip with 1.2 M sensors generates approximately 25 million bases. The performance of the ion chips and overall sequencing platform is demonstrated through whole-genome sequencing of three bacterial genomes. The scalability of our chip architecture is demonstrated by producing chips with up to 10 times the number of sensors and producing a low-coverage sequence of the genome of Gordon Moore, author of Moore's law<sup>16</sup>.

## A CMOS integrated circuit for sequencing

We have developed a simple, scalable ISFET sensor architecture using electronic addressing common in modern CMOS imagers (Supplementary Fig. 2). Our integrated circuit consists of a large array of sensor elements, each with a single floating gate connected to an underlying ISFET (Fig. 1a). For sequence confinement we rely on a

<sup>1</sup>Ion Torrent by Life Technologies, Suite 100, 246 Goose Lane, Guilford, Connecticut 06437, USA.



**Figure 1 | Sensor, well and chip architecture.** **a**, A simplified drawing of a well, a bead containing DNA template, and the underlying sensor and electronics. Protons ( $H^+$ ) are released when nucleotides (dNTP) are incorporated on the growing DNA strands, changing the pH of the well ( $\Delta pH$ ). This induces a change in surface potential of the metal-oxide-sensing layer, and a change in potential ( $\Delta V$ ) of the source terminal of the underlying field-effect

transistor. **b**, Electron micrograph showing alignment of the wells over the ISFET metal sensor plate and the underlying electronic layers. **c**, Sensors are arranged in a two-dimensional array. A row select register enables one row of sensors at a time, causing each sensor to drive its source voltage onto a column. A column select register selects one of the columns for output to external electronics.

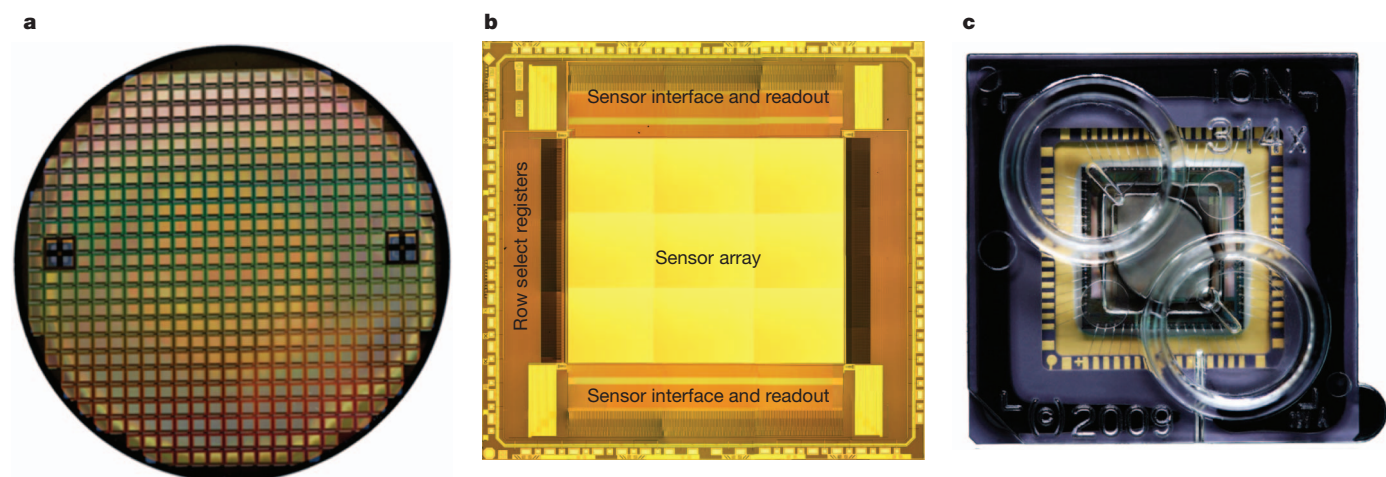
3.5- $\mu m$ -diameter well formed by adding a 3- $\mu m$ -thick dielectric layer over the electronics and etching to the sensor plate (Fig. 1b). A tantalum oxide layer provides for proton sensitivity (58 mV  $pH^{-1}$ ; ref. 38). High-speed addressing and readout are accomplished by the semiconductor electronics integrated with the sensor array (Fig. 1c). The sensor and underlying electronics provide a direct transduction from the incorporation event to an electronic signal. Unlike light-based sequencing technology, we do not use the elements of the array to collect photons and form a larger image to detect the incorporation of a base; instead we use each sensor to independently and directly monitor the hydrogen ions released during nucleotide incorporation.

Ion chips are manufactured on wafers (Fig. 2a), cut into individual die (Fig. 2b) and robotically packaged with a disposable polycarbonate flow cell that isolates the fluids to regions above the sensor array and away from the supporting electronics to provide convenient sample loading as well as electrical and fluidic interfaces to the sequencing instrument (Fig. 2c). Chips were designed and fabricated with 1.5 M, 7.2 M and 13 M ISFETs (Supplementary Fig. 3). On the basis of the placement of the flow cell on the sensor array, 1.2 M, 6.1 M and 11 M

wells and sensors are exposed to fluids, with 99.9% of the sensors sensitive to pH and usable for DNA sequencing (Supplementary Fig. 4). Increasing the numbers of sensors per chip was first achieved by increasing the die area, from 10.6 mm  $\times$  10.9 mm to 17.5 mm  $\times$  17.5 mm, and then by increasing the density of the sensors by reducing the number of transistors per sensor from three to two. Chip density is limited by the selection of the CMOS node and the number of transistors per sensing element. Using a 0.35  $\mu m$  CMOS node the minimum spacing for a three-transistor sensor is 5.1  $\mu m$  and for a two-transistor sensor it is 3.8  $\mu m$  (Supplementary Fig. 5). To understand further the limits on density, we show that 1.3  $\mu m$  wells are readily manufactured, can be aligned to sensors, enable the generation of high-quality sequence (Supplementary Fig. 6) and can, using a 110 nm node, be fabricated with a spacing as small as 1.68  $\mu m$  (Supplementary Fig. 7).

### Sequencing on a semiconductor device

The all-electronic detection system used by the ion chip simplifies and greatly reduces the cost of the sequencing instrument (Supplementary



**Figure 2 | Wafer, die and chip packaging.** **a**, Fabricated CMOS 8'' wafer containing approximately 200 individual functional ion sensor die. **b**, Unpackaged die, after automated dicing of wafer, with functional regions

indicated. **c**, Die in ceramic package wire bonded for electrical connection, shown with moulded fluidic lid to allow addition of sequencing reagents.



Fig. 8). The instrument has no optical components, and is comprised primarily of an electronic reader board to interface with the chip, a microprocessor for signal processing, and a fluidics system to control the flow of reagents over the chip (Supplementary Fig. 9).

Genomic DNA is prepared for sequencing as described in Supplementary Methods. Briefly, DNA is fragmented, ligated to adapters, and adaptor-ligated libraries are clonally amplified onto beads. Template-bearing beads are enriched through a magnetic-bead-based process. Sequencing primers and DNA polymerase are then bound to the templates and pipetted into the chip's loading port. Individual beads are loaded into individual sensor wells by spinning the chip in a desktop centrifuge. A 2  $\mu\text{m}$  acrylamide bead was chosen to deliver sufficient copies of the template to the sensor well to achieve a high signal-to-noise ratio (SNR) (800 K copies, SNR, 10; Supplementary Methods and Supplementary Fig. 10), while well depth was selected to allow only a single bead to occupy a well.

In ion sequencing, all four nucleotides are provided in a stepwise fashion during an automated run (Supplementary Methods). When the nucleotide in the flow is complementary to the template base directly downstream of the sequencing primer, the nucleotide is incorporated into the nascent strand by the bound polymerase. This increases the length of the sequencing primer by one base (or more, if a homopolymer stretch is directly downstream of the primer) and results in the hydrolysis of the incoming nucleotide triphosphate, which causes the net liberation of a single proton for each nucleotide incorporated during that flow. The release of the proton produces a shift in the pH of the surrounding solution proportional to the number of nucleotides incorporated in the flow (0.02 pH units per single base incorporation). This is detected by the sensor on the bottom of each well, converted to a voltage and digitized by off-chip electronics (Fig. 3). The signal generation and detection occurs over 4 s (Fig. 3b). After the flow of each nucleotide, a wash is used to ensure nucleotides do not remain in the well. The small size of the wells allows diffusion into and out of the well on the order of a one-tenth of a second and eliminates the need for enzymatic removal of reagents<sup>1</sup>.

### Signal processing and base calling

To change raw voltages into base calls, signal-processing software converts the raw data into measurements of incorporation in each well for each successive nucleotide flow using a physical model. Sampling the signal at high frequency relative to the time of the incorporation signal allows signal averaging to improve the SNR. The physical model takes into consideration diffusion rates, buffering effects and polymerase rates (Supplementary Fig. 11). The model is

applied and fit to the raw trace from each well and the incorporation signals are extracted. A base caller corrects the signals for phase and signal loss, normalizes to the key, and generates corrected base calls for each flow in each well to produce the sequencing reads (Fig. 3c and Supplementary Fig. 12).

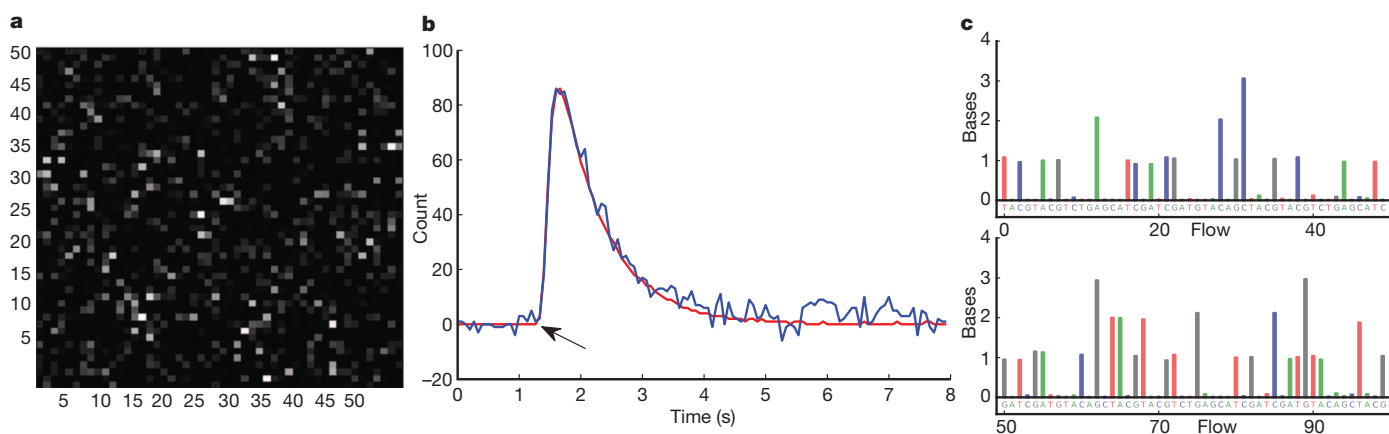
Next, each read is sequentially passed through two signal-based filters to exclude low-accuracy reads. The first filter measures the fraction of flows in which an incorporation event was measured. When this value is unusually large (greater than 60% of the first 60 flows) the read is not clonal. The second filter measures the extent to which the observed signal values match those predicted by the phasing model. When there is poor agreement (median absolute difference more than 0.06 over the first 60 flows) between the two, it corresponds to higher error rates. Lastly, per-base quality values are predicted using an adaptation of the Phred method<sup>39</sup> that quantifies the concordance between the phasing model predictions and the observed signal. These *ab initio* scores track closely with post-alignment derived quality scores, and are used to trim back low-quality sequence from the 3' end of a read (Supplementary Fig. 13).

### Sequencing bacterial genomes

Bacterial genome sequencing and signal processing was performed as described earlier. We succeeded in sequencing all three genomes five-fold to tenfold in individual runs using the small ion chip, covering 96.80% to 99.99% of each genome, with genome-wide consensus accuracies as high as 99.99% (Table 1 and Supplementary Fig. 14). *Escherichia coli* sequencing with three successively larger ion chips produced 46 to over 270 megabases of sequence (Table 1).

To characterize run quality, we aligned each read to the corresponding reference genome (Supplementary Fig. 15). The per-base accuracy was observed to be  $99.569\% \pm 0.001\%$  within the first 50 bases and  $98.897\% \pm 0.001\%$  within the first 100 bases (Supplementary Fig. 16a). This accuracy is similar at 50 bases and higher at 100 bases than light-based methods using modified nucleotides (1.1% versus 5% error<sup>40</sup>). The per-base accuracy in calling a homopolymer of length 5 is  $97.328\% \pm 0.023\%$  (Supplementary Fig. 16b) and higher than pyrosequencing-based sequencing methods<sup>1,41</sup>. For each genome, the observed distribution of per-base coverage matches closely with the theoretical Poisson distribution reflecting the uniform nature of the coverage (Supplementary Fig. 17). The distribution of coverage was also relatively unbiased across GC content (Supplementary Fig. 18).

Ion sequencing technology has allowed the routine acquisition of 100-base read lengths, and perfect read lengths exceeding 200 bases (Supplementary Fig. 19). At present, 20–40% of the sensors in a given



**Figure 3 | Data collection and base calling.** **a**, A  $50 \times 50$  region of the ion chip. The brightness represents the intensity of the incorporation reaction in individual sensor wells. **b**, 1-nucleotide incorporation signal from an individual sensor well; the arrow indicates start of incorporation event, with the physical

model (red line) and background corrected data (blue line) shown. **c**, The first 100 flows from one well. Each coloured bar indicates the corresponding number of bases incorporated during that nucleotide flow.

**Table 1** | *Vibrio fischeri*, *E. coli*, *Rhodopseudomonas palustris* and *Homo sapiens*

	<i>V. fischeri</i>	<i>R. palustris</i>	<i>E. coli</i>	<i>E. coli</i>	<i>E. coli</i>	<i>H. sapiens</i>
GC content	38%	65%	51%	51%	51%	41%
Genome size	4.2 Mb	5.5 Mb	4.7 Mb	4.7 Mb	4.7 Mb	2.9 Gb
Number of runs x ion chip size	1 × 1.2 M	1 × 1.2 M	1 × 1.2 M	1 × 6.1 M	1 × 11 M	1,601 × 1.2 M 267 × 6.1 M 28 × 11.1 M
Fold coverage	6.2-fold	6.9-fold	11.3-fold	36.2-fold	58.4-fold	10.6-fold
Coverage	96.80%	99.64%	99.99%	100.00%	100.00%	99.21%
Reads ≥21 bases	261,313	444,750	507,198	1,852,931	2,594,031	366,623,578
Reads ≥50 bases	233,049	399,360	487,420	1,698,852	2,343,880	306,042,650
Reads ≥100 bases	156,391	160,726	400,743	1,012,918	1,779,237	139,624,090
Mapped bases	26.0 Mb	37.8 Mb	47.6 Mb	169.6 Mb	273.9 Mb	30.2 Gb

Coverage shows percentage of genome covered based on one or more reads mapping to each base of the reference genome. Reads align with 98% or greater accuracy.

run yield mappable reads. The gap between the number of sensors on a chip and the number yielding sequence is primarily the result of incomplete loading of the chip, poor amplification of a fragment onto the bead, and lack of clonality of the template. With continued improvements in loading and template preparation, along with improvements in signal processing and base calling, it is expected that the percentage of sensors yielding reads, the average read length and read accuracy will all improve significantly, as it has for other sequencing technologies<sup>1–4,9–11</sup>.

### 'Post-light' sequencing of G. Moore

To illustrate the scalability of semiconductor sequencing we produced whole-genome sequence data from an individual, G. Moore<sup>42</sup> (Fig. 4). Written consent was provided by G. Moore to sequence and publish his genome and resulting findings. Reads from his genome were deposited in the Sequence Read Archive (SRA) under accession number ERP000682. The mean coverage of the G. Moore genome was 10.6-fold (Table 1). The degree to which the observed distribution of reads conforms to a Poisson distribution is indicative of a general lack of bias in coverage depth (Fig. 4b).

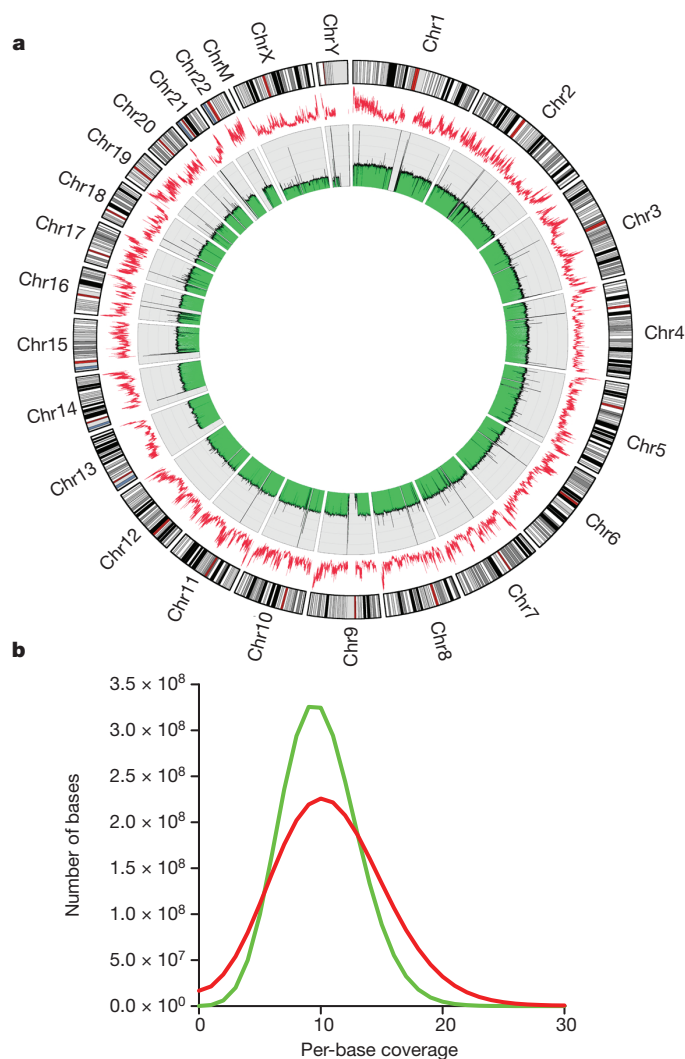
We found 2,598,983 SNPs in the G. Moore genome, of which 3.08% were found to be novel, consistent with previous reports<sup>4,9,11</sup> (Supplementary Methods). To confirm the accuracy of our analysis, we also sequenced the G. Moore genome using ABI SOLiD Sequencing<sup>43</sup> to 15-fold coverage and validated 99.95% of the heterozygous and 99.97% of the homozygous genotypes (Supplementary Tables 1 and 2).

We used the Online Mendelian Inheritance in Man database<sup>44</sup> and the 23andMe functional SNP collection (<https://www.23andme.com>) to identify a subset of validated SNPs known to be involved in human disease and interesting phenotypes (Supplementary Table 3). We also examined the G. Moore sequence for the 7,693 deletions and inversions discovered by the 1000 Genomes Consortium and computationally found 3,413 of them in the G. Moore genome at a 99.94% positive predictive value (Supplementary Methods, Supplementary Table 4 and Supplementary Fig. 20). To determine G. Moore's maternal ancestry, reads were also mapped to human mitochondrial DNA<sup>45</sup> for a mean coverage of 732-fold. G. Moore's mitochondria belong to haplogroup H, the most common in Europe<sup>46</sup>.

### Discussion

We have demonstrated the ability to produce and use a disposable integrated circuit fabricated in standard CMOS foundries to perform, for the first time, 'post-light' genome sequencing of bacterial and human genomes. With fifty billion dollars spent per year on CMOS semiconductor fabrication and packaging technologies, our goal was to leverage that investment to make a highly scalable sequencing technology. Using the G. Moore genome we demonstrated the feasibility of sequencing a human genome. The G. Moore genome sequence required on the order of a thousand individual ion chips

comprising about one billion sensors. By demonstrating the ability to make larger and denser arrays, use fewer transistors per sensor, and sequence from wells as small as 1.3  $\mu\text{m}$ , our work suggests that readily available CMOS nodes should enable the production of one-billion-sensor ion chips and low-cost routine human genome sequencing.



**Figure 4** | **G. Moore genome.** **a**, Circular genome plot. The average sequencing coverage (green) and average GC content (red) within 100-kb intervals is shown. **b**, Distribution of the observed per-base coverage depth along the genome (red) compared with the distribution expected from random coverage (green).



Received 8 March; accepted 26 May 2011.

1. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
2. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
3. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
4. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
5. Thomas, R. K. *et al.* Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nature Med.* **12**, 852–855 (2006).
6. Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
7. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genet.* **42**, 30–35 (2010).
8. Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**, 223–227 (2005).
9. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
10. Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
11. Schuster, S. C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463**, 943–947 (2010).
12. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
13. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl Acad. Sci. USA* **103**, 12115–12120 (2006).
14. Noonan, J. P. *et al.* Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597–599 (2005).
15. Green, R. E. *et al.* Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330–336 (2006).
16. Moore, G. E. Cramming more components onto integrated circuits. *Electronics* **38**, 114–117 (1965).
17. Davies, K. *The \$1,000 Genome* (Free Press, 2010).
18. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
19. Smith, L. M. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986).
20. Metzker, M. L. Sequencing technologies—the next generation. *Nature Rev. Genet.* **11**, 31–46 (2010).
21. Wanlass, F. M. Low stand-by power complementary field effect circuitry. U.S. patent 3,356,858: (1967).
22. Theuvsen, A. J. P. CMOS image sensors: state-of-the-art. *Solid-State Electron.* **52**, 1401–1406 (2008).
23. Sakurai, T. & Husimi, Y. Real-time monitoring of DNA polymerase reactions by a micro ISFET pH sensor. *Anal. Chem.* **64**, 1996–1997 (1992).
24. Fritz, J., Cooper, E. B., Gaudet, S., Sorger, P. K. & Manalis, S. R. Electronic detection of DNA by its intrinsic molecular charge. *Proc. Natl Acad. Sci. USA* **99**, 14142–14146 (2002).
25. Drummond, T. G., Hill, M. G. & Barton, J. K. Electrochemical DNA sensors. *Nature Biotechnol.* **21**, 1192–1199 (2003).
26. Bergveld, P. Development of an ion-sensitive solid-state device for neurophysiological measurements. *IEEE Trans. Biomed. Eng.* **17**, 70–71 (1970).
27. Bergveld, P. Thirty years of ISFETOLOGY. What happened in the past 30 years and what may happen in the next 30 years. *Sens. Actuators B Chem.* **88**, 1–20 (2003).
28. Yeow, T., Haskard, M., Mulcahy, D., Seo, H. & Kwon, D. A very large integrated pH-ISFET sensor array chip compatible with standard CMOS processes. *Sens. Actuators B Chem.* **44**, 434–440 (1997).
29. Bausells, J., Carrabina, J., Errachid, A. & Merlos, A. Ion-sensitive field-effect transistors fabricated in a commercial CMOS technology. *Sens. Actuators B Chem.* **57**, 56–62 (1999).
30. Milgrew, M., Hammond, P. & Cumming, D. The development of scalable sensor arrays using standard CMOS technology. *Sens. Actuators B Chem.* **103**, 37–42 (2004).
31. Hizawa, T., Sawada, K., Takao, H. & Ishida, M. Fabrication of a two-dimensional pH image sensor using a charge transfer technique. *Sens. Actuators B Chem.* **117**, 509–515 (2006).
32. Purushothaman, S., Toumazou, C. & Ou, C. P. Protons and single nucleotide polymorphism detection: A simple use for the ion sensitive field effect transistor. *Sens. Actuators B Chem.* **114**, 964–968 (2006).
33. Pourmand, N. *et al.* Direct electrical detection of DNA synthesis. *Proc. Natl Acad. Sci. USA* **103**, 6466–6470 (2006).
34. Sakata, T. & Miyahara, Y. DNA sequencing based on intrinsic molecular charges. *Angew. Chem.* **118**, 2283–2286 (2006).
35. Milgrew, M. J., Riehle, M. O. & Cumming, D. R. S. A large transistor-based sensor array chip for direct extracellular imaging. *Sens. Actuators B Chem.* **111–112**, 347–353 (2005).
36. Milgrew, M. J. & Cumming, D. R. S. Matching the transconductance characteristics of CMOS ISFET arrays by removing trapped charge. *Electron Devices. IEEE Trans. Electron Devices* **55**, 1074–1079 (2008).
37. Hammond, P. & Cumming, D. Encapsulation of a liquid-sensing microchip using SU-8 photoresist. *Microelectron. Eng.* **73–74**, 893–897 (2004).
38. Mikolajick, T., Kühnhold, R. & Rysse, H. The pH-sensing properties of tantalum pentoxide films fabricated by metal organic low pressure chemical vapor deposition. *Sens. Actuators B Chem.* **44**, 262–267 (1997).
39. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
40. Claesson, M. J. *et al.* Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* **38**, e200 (2010).
41. Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L. & Mark Welch, D. Accuracy and quality of massively-parallel DNA pyrosequencing. *Genome Biol.* **8**, R143 (2007).
42. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
43. McKernan, K. J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541 (2009).
44. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517 (2005).
45. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genet.* **23**, 147 (1999).
46. Kloss Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We want to thank G. Moore for his willingness to participate in this study. We thank G. Fergus, M. Jain, J. Kole, L. Stevens and the ION team for supporting our efforts, and H. Peckman, V. Tadigotla, D. Holloway and S. McLaughlin for help on the variant analysis, and M. Ross of the Broad Institute for help on quality scores. This research was supported, in part, by a grant from the National Human Genome Research Institute (NHGRI), RFA-HG-08-008, Revolutionary Genome Sequencing Technologies—The \$1000 Genome. Grant number: R01 HG005094.

**Author Contributions** J.M.R. conceived the technology, supervised the project and wrote the manuscript with input from co-authors. K.J., M.J.M. and J.B. designed chips. J.F.D., M.A., D.L., J.W.M., J.F.S., E.N., M.S., X.M., A.B., T.A.C., M.H., I.B.S., B.R., J.S., E.F., M.S., J.A.F., K.J.M. and J.H.L. developed methods. M.D., J.T.B., M.E., J.H., N.H., T.M.R., B.P.P., S.E.C., M.L., Y.F. and A.W. wrote software and analysed data. W.H., J.S., W.M., D.M., J.R.N. and G.T.R. designed the instrument. E.D., D.D., R.K. and T.S. sequenced the human sample.

**Author Information** Sequences for *Homo sapiens*, *Escherichia coli*, *Vibrio fischeri* and *Rhodospseudomonas palustris* were deposited in the Sequence Read Archive (SRA) under accession numbers ERP000682, ERP000541, ERP000542 and ERP000543 respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike Licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare competing financial interests: details accompany the full-text HTML version of this paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.M.R. (Jonathan.Rothberg@LifeTech.com).

# Sideband cooling of micromechanical motion to the quantum ground state

J. D. Teufel<sup>1</sup>, T. Donner<sup>2,3</sup>, Dale Li<sup>1</sup>, J. W. Harlow<sup>2,3</sup>, M. S. Allman<sup>1,3</sup>, K. Cicak<sup>1</sup>, A. J. Sirois<sup>1,3</sup>, J. D. Whittaker<sup>1,3</sup>, K. W. Lehnert<sup>2,3</sup> & R. W. Simmonds<sup>1</sup>

The advent of laser cooling techniques revolutionized the study of many atomic-scale systems, fuelling progress towards quantum computing with trapped ions<sup>1</sup> and generating new states of matter with Bose–Einstein condensates<sup>2</sup>. Analogous cooling techniques<sup>3,4</sup> can provide a general and flexible method of preparing macroscopic objects in their motional ground state. Cavity optomechanical or electromechanical systems achieve sideband cooling through the strong interaction between light and motion<sup>5–15</sup>. However, entering the quantum regime—in which a system has less than a single quantum of motion—has been difficult because sideband cooling has not sufficiently overwhelmed the coupling of low-frequency mechanical systems to their hot environments. Here we demonstrate sideband cooling of an approximately 10-MHz micromechanical oscillator to the quantum ground state. This achievement required a large electromechanical interaction, which was obtained by embedding a micromechanical membrane into a superconducting microwave resonant circuit. To verify the cooling of the membrane motion to a phonon occupation of  $0.34 \pm 0.05$  phonons, we perform a near-Heisenberg-limited position measurement<sup>3</sup> within  $(5.1 \pm 0.4)\hbar/2\pi$ , where  $\hbar$  is Planck's constant. Furthermore, our device exhibits strong coupling, allowing coherent exchange of microwave photons and mechanical phonons<sup>16</sup>. Simultaneously achieving strong coupling, ground state preparation and efficient measurement sets the stage for rapid advances in the control and detection of non-classical states of motion<sup>17,18</sup>, possibly even testing quantum theory itself in the unexplored region of larger size and mass<sup>19</sup>. Because mechanical oscillators can couple to light of any frequency, they could also serve as a unique intermediary for transferring quantum information between microwave and optical domains<sup>20</sup>.

A mechanical oscillator of high quality factor placed within the quantum regime could allow us to explore quantum mechanics in entirely new ways<sup>17–20</sup>. To do this requires the ability to prepare the oscillator in its ground state, to arbitrarily control its quantum states, and to detect these states near the Heisenberg limit. In addition, the oscillator system should not be strongly perturbed by its environment or any other extraneous influence, including dissipation or thermal excitations. As a first step, the oscillator's temperature  $T$  must be reduced so that  $k_B T < \hbar\Omega_m$ , where  $\Omega_m$  is the resonance frequency of the oscillator,  $k_B$  is Boltzmann's constant, and  $\hbar$  is  $\hbar/2\pi$ . Although there has been substantial progress in cooling mechanical oscillators with radiation pressure forces, sideband cooling to the quantum mechanical ground state has been a long-standing challenge. Cavity optomechanical systems have realized very large sideband cooling rates<sup>9–13,15</sup>; however, these rates are not sufficient to overcome the larger thermal heating rates of the mechanical modes. Electromechanical experiments using much lower-energy microwave photons<sup>5–8,14</sup>, although simpler to operate below 100 mK, have suffered from weak electromechanical interactions and inefficient detection of the photon fields.

In a unique approach, a system based on a high-frequency (6-GHz) microwave dilatation oscillator was integrated with a superconducting

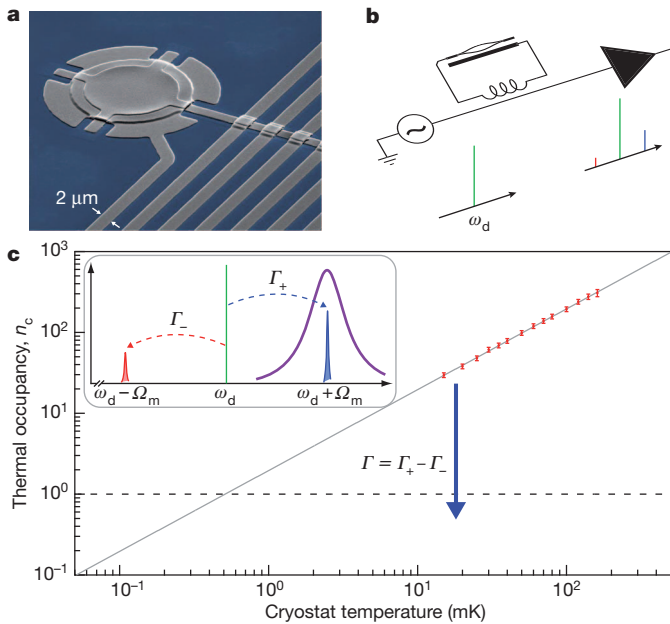
phase qubit<sup>21</sup>. Its high frequency offered the advantage of reaching the ground state at relatively high temperatures ( $T \approx 25$  mK), which were achievable simply with passive dilution refrigeration. Furthermore, because the mechanical oscillator was piezoelectric, its strong electrical response enabled significantly strong coupling to the superconducting qubit, providing direct control and (destructive) measurement of the phonon energy states. These results showed that quantum effects are achievable with a human-fabricated mechanical oscillator. Unfortunately, the short mechanical lifetimes prevented the manipulation of complex mechanical states and direct tests of entanglement.

Low frequency (<100-MHz) mechanical oscillators have distinct advantages: higher quality factors, long phonon lifetimes and large motional state displacements, which are important for future tests of quantum theory<sup>19</sup>. Cavity opto- or electro-mechanical systems<sup>4</sup> naturally offer a powerful method for both cooling and detecting low-frequency mechanical oscillators<sup>22,23</sup>. An object whose motion alters the resonance frequency,  $\omega_c$ , of an electromagnetic cavity experiences a radiation pressure force governed by the parametric interaction Hamiltonian:  $\hat{H}_{\text{int}} = \hbar G \hat{n} \hat{x}$ , where  $G = d\omega_c/dx$ ,  $\hat{n}$  is the cavity photon number, and  $\hat{x}$  is the displacement of the mechanical oscillator. By driving the cavity at a frequency  $\omega_d$ , the oscillator's motion produces upper and lower sidebands at  $\omega_d \pm \Omega_m$ . Because these sideband photons are inelastically scattered from the drive field, they provide a way to exchange energy with the oscillator. If the drive field is optimally detuned below the cavity resonance by an amount  $\Delta \equiv \omega_d - \omega_c = -\Omega_m$ , photons will be preferentially up-converted to  $\omega_c$  because the photon density of states is maximal there (Fig. 1b). When an up-converted photon leaves the cavity, it removes the energy of one mechanical quantum (one phonon) from the motion. Thus, the mechanical oscillator is damped and cooled by way of this radiation-pressure force. Because the mechanical motion is encoded in the scattered photons leaving the cavity, information on the position of the mechanical oscillator provides a near Heisenberg-limited measurement of displacement<sup>24</sup>.

Here we use a cavity electromechanical system where a flexural mode of a thin aluminium membrane is parametrically coupled to a superconducting microwave resonant circuit. Unlike previous microwave systems, this device achieves large electromechanical coupling by use of a flexible vacuum-gap capacitor<sup>16,25</sup>. The oscillator is a 100-nm-thick aluminium membrane with a diameter of 15  $\mu\text{m}$ , suspended 50 nm above a second aluminium layer on a sapphire substrate<sup>25</sup> (Fig. 1). These two metal layers form a parallel-plate capacitor that is shunted by a 12-nH spiral inductor. This combination of capacitor and inductor creates a microwave cavity with a displacement-dependent resonance frequency centred at  $\omega_c = 2\pi \times 7.54$  GHz. The device is operated in a dilution refrigerator at 15 mK, at which temperature aluminium is superconducting, and the microwave cavity has a total energy decay rate of  $\kappa \approx 2\pi \times 200$  kHz. The diameter of the aluminium membrane and its tension<sup>8</sup> produce an  $\Omega_m$  of  $2\pi \times 10.56$  MHz with an intrinsic damping rate of  $\Gamma_m = 2\pi \times 32$  Hz, or mechanical quality factor  $Q_m = \Omega_m/\Gamma_m = 3.3 \times 10^5$ . The oscillator mass,  $m = 48$  pg, implies that

<sup>1</sup>National Institute of Standards and Technology (NIST), Boulder, Colorado 80305, USA. <sup>2</sup>JILA, University of Colorado and NIST, Boulder, Colorado 80309, USA. <sup>3</sup>Department of Physics, University of Colorado, Boulder, Colorado 80309, USA.





**Figure 1 | Schematic description of the experiment.** **a**, False-colour scanning electron micrograph showing the aluminium (grey) electromechanical circuit fabricated on a sapphire (blue) substrate; a 15-μm-diameter membrane is suspended 50 nm above a lower electrode. The membrane's motion modulates the capacitance, and hence, the resonance frequency of the superconducting microwave circuit. **b**, A coherent microwave drive (left,  $\omega_d$ , shown green in frequency-amplitude plot below) inductively coupled to the circuit (top) acquires modulation sidebands (red and blue in plot below) owing to the thermal motion of the membrane. The upper sideband is amplified with a nearly quantum-limited Josephson parametric amplifier (filled triangle, right) within the cryostat. **c**, The microwave power in the upper sideband provides a direct measurement of the thermal occupancy of the mechanical mode, which may be calibrated *in situ* by varying the temperature of the cryostat (main panel). The mechanical mode shows thermalization with the cryostat at all temperatures, yielding a minimum thermal occupancy of 30 mechanical quanta without using sideband-cooling techniques. Error bars, s.d. Inset, illustration of the concept of sideband cooling. When the circuit is excited with a detuned microwave drive such that  $\Delta = -\Omega_m$ , the narrow line shape of the electrical resonance ensures that the rate to scatter photons to higher energy  $\Gamma_+$  (blue dashed arrow, blue peak) exceeds the rate to scatter to lower energy  $\Gamma_-$  (red dashed arrow, red peak). Thus, the net scattering rate  $\Gamma$  (blue solid arrow) provides a cooling mechanism for the membrane.

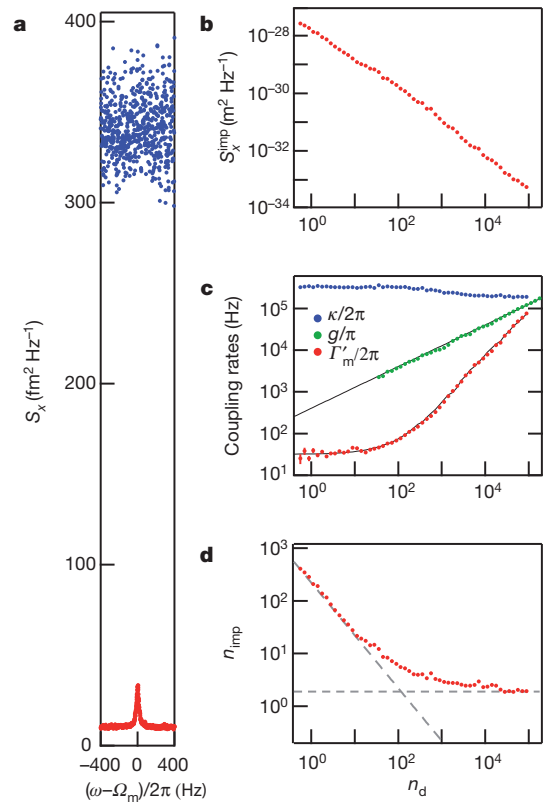
the zero-point motion is  $x_{zp} = \sqrt{\hbar/(2m\Omega_m)} = 4.1$  fm. With a ratio of  $\Omega_m/\kappa > 50$ , our system is deep within the resolved-sideband regime and well-suited for sideband cooling to the mechanical ground state<sup>22,23</sup>.

To measure the mechanical displacement, we apply a microwave field, which is detuned below the cavity resonance frequency by  $\Delta = -\Omega_m$ , through heavily attenuated coaxial lines to the feed line of our device. The upper sideband, now at  $\omega_c$ , is amplified with a custom-built Josephson parametric amplifier<sup>26,27</sup> followed by a low-noise cryogenic amplifier, demodulated at room temperature, and finally monitored with a spectrum analyser. The thermal motion of the membrane creates an easily resolvable peak in the microwave noise spectrum. As described previously<sup>27</sup>, this measurement scheme constitutes a nearly shot-noise-limited microwave interferometer with which we can measure mechanical displacement with minimum added noise close to fundamental limits.

In order to calibrate the demodulated signal to the membrane's motion, we measure the thermal noise spectrum while varying the cryostat temperature (Fig. 1c). Here a weak microwave drive ( $\sim 3$  photons in the cavity) is used in order to ensure that radiation pressure damping and cooling effects are negligible. When  $\Omega_m \gg \kappa \gg \Gamma_m$  and  $\Delta = -\Omega_m$ , the displacement spectral density  $S_x$  is related to the observed microwave noise spectral density  $S$  by  $S_x = 2(\kappa\Omega_m/G\kappa_{ex})^2 S/P_o$ , where  $\kappa_{ex}$  is the

coupling rate between the cavity and the feed line, and  $P_o$  is the power of the microwave drive at the output of the cavity. According to equipartition, the area under the resonance curve of displacement spectral density  $S_x$  must be proportional to the effective temperature of the mechanical mode. This calibration procedure allows us to convert the sideband in the microwave power spectral density to a displacement spectral density and to extract the thermal occupation of the mechanical mode. In Fig. 1c we show the number of thermal quanta in the mechanical resonator as a function of  $T$ . The linear dependence of the integrated power spectral density with temperature shows that the mechanical mode equilibrates with the cryostat even for the lowest achievable temperature of 15 mK. This temperature corresponds to a thermal occupancy  $n_m = 30$ , where  $n_m = [\exp(\hbar\Omega_m/k_B T) - 1]^{-1}$ . The calibration determines the electromechanical coupling strength,  $G/2\pi = 49 \pm 2$  MHz nm<sup>-1</sup>. With these device parameters, we can investigate both the fundamental sensitivity of our measurement and the effects of radiation pressure cooling.

The total measured displacement noise results from two sources: the membrane's actual mean-square motion,  $S_x^{\text{th}}$ , and its apparent motion,  $S_x^{\text{imp}}$ , which is due to imprecision of the measurement. Figure 2a demonstrates how the use of low-noise parametric amplification significantly



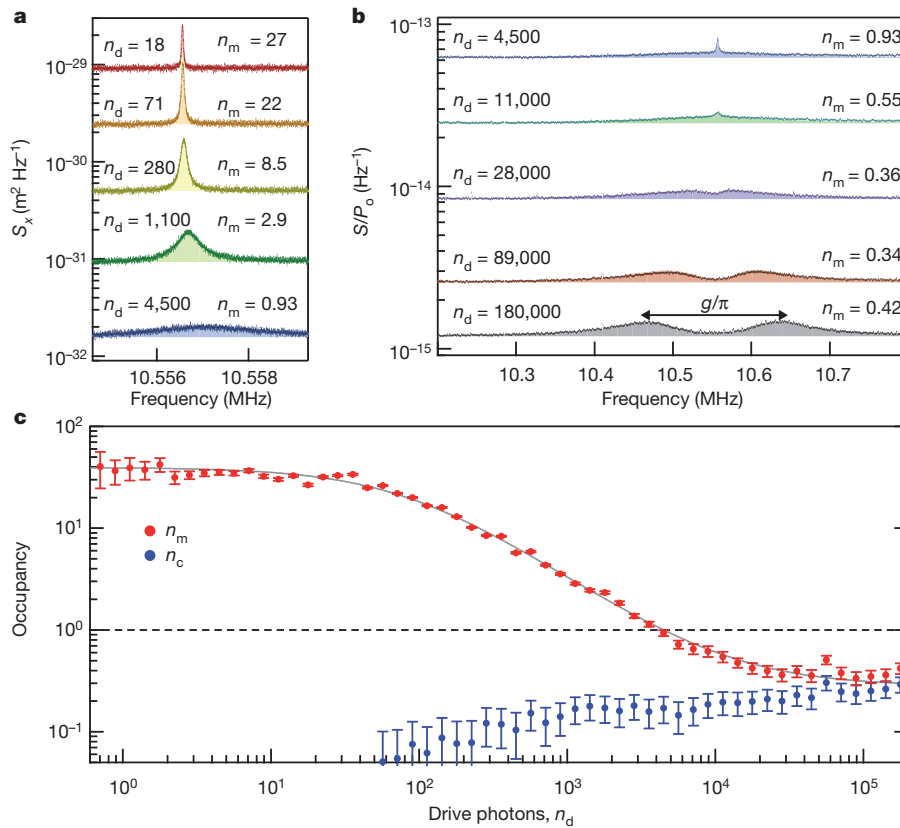
**Figure 2 | Displacement sensitivity in the presence of dynamical back-action.** **a**, The displacement spectral density  $S_x$  measured with (red) and without (blue) the Josephson parametric amplifier. As the parametric amplifier greatly reduces the total noise of the microwave measurement, the time required to resolve the thermal motion is reduced by a factor of 1,000. **b**, As the microwave drive power is increased, the absolute displacement sensitivity,  $S_x^{\text{imp}}$  improves, reaching a minimum of  $5.5 \times 10^{-34}$  m<sup>2</sup> Hz<sup>-1</sup> at the highest power. **c**, The parametric coupling rate  $g$  between the microwave cavity and the mechanical mode increases as  $\sqrt{n_d}$ . This coupling broadens the linewidth of the mechanical mode  $\Gamma_m$  from its intrinsic value of  $\Gamma_m = 2\pi \times 32$  Hz until it exceeds the linewidth of the microwave cavity  $\kappa$ . **d**, The relative measurement imprecision, in units of mechanical quanta, depends on the product of  $S_x^{\text{imp}}$  and  $\Gamma_m$ . Thus, once the power is large enough that dynamical back-action overwhelms the intrinsic mechanical linewidth,  $n_{\text{imp}}$  asymptotically approaches a constant value ( $n_{\text{imp}} = 1.9$ ), which is a direct measure of the overall efficiency of the photon measurement.

lowers  $S_x^{\text{imp}}$ , resulting in a reduction in the white-noise background by a factor of more than 30. This greatly increases the signal-to-noise ratio of the membrane's thermal motion, thereby reducing the integration time required to resolve the thermal peak by a factor of 1,000. To investigate the measurement sensitivity in the presence of dynamical back-action, we regulate the cryostat temperature at 20 mK and increase the amplitude of the detuned microwave drive while observing modifications in the displacement spectral density. We quantify the strength of the drive by the resulting number of photons  $n_d$  in the microwave cavity. As shown in Fig. 2b, the measurement imprecision  $S_x^{\text{imp}}$  is inversely proportional to  $n_d$ . At the highest drive power ( $n_d \approx 10^5$ ), the absolute displacement sensitivity is  $5.5 \times 10^{-34} \text{ m}^2 \text{ Hz}^{-1}$ .

As expected, the increased drive power also damps and cools the mechanical oscillator<sup>3,22,23</sup>. The total mechanical dissipation rate  $\Gamma'_m = \Gamma_m + \Gamma$  is the sum of the intrinsic dissipation,  $\Gamma_m$ , and the radiation-pressure-induced damping resulting from scattering photons to the upper/lower sideband,  $\Gamma = \Gamma_+ - \Gamma_-$ , where  $\Gamma_{\pm} = 4g^2\kappa/[\kappa^2 + 4(\Delta \pm \Omega_m)^2]$ . Here  $g$  is the coupling rate between the cavity and the mechanical mode, which depends on the amplitude of the drive:  $g = Gx_{\text{zp}}\sqrt{n_d}$ . Figure 2c shows the measured values of  $\kappa$ ,  $g$  and  $\Gamma'_m$  as the drive increases. The radiation-pressure damping of the mechanical oscillator becomes pronounced above a cavity drive amplitude of approximately 75 photons, at which point  $\Gamma'_m = 2\Gamma_m$  and the mechanical linewidth has doubled. Note that the increased damping rate can be switched off at any time by removing the cooling drive, returning the mechanical oscillator to its intrinsic quality factor,  $Q_m$ .

Whereas the absolute value of the displacement imprecision decreases with increasing power, the visibility of the thermal mechanical peak no longer improves once the radiation-pressure force becomes the dominant dissipation mechanism for the membrane. By expressing the imprecision as equivalent thermal quanta of the oscillator,  $n_{\text{imp}} = \Gamma'_m S_x^{\text{imp}} / 8x_{\text{zp}}^2$ , we see that the visibility of the thermal noise above the imprecision no longer improves once the drive is much greater than  $n_d \approx 100$  (Fig. 2d). This is because a linear decrease in  $S_x^{\text{imp}}$  is balanced by a linear increase in  $\Gamma'_m$  due to radiation-pressure damping. The asymptotic value of  $n_{\text{imp}}$  is a direct measure of the efficiency of the microwave measurement. Ideally, for a lossless circuit, a quantum-limited microwave measurement would imply  $n_{\text{imp}} = 1/4$ . The incorporation of the low-noise Josephson parametric amplifier improves  $n_{\text{imp}}$  close to this ideal limit, reducing the asymptotic value of  $n_{\text{imp}}$  from 70 to 1.9 quanta. This level of sensitivity is crucial for resolving any residual thermal motion when cooling into the quantum regime.

Beginning from a cryostat temperature of 20 mK and a thermal occupation of  $n_m^T = 40$  quanta, the fundamental mechanical mode of the membrane is cooled by radiation-pressure forces. Figure 3a shows the displacement spectral density of the motional sideband as  $n_d$  is increased from 18 to 4,500 photons, along with fits to a Lorentzian lineshape (shaded areas). As described above, this increased drive results in three effects on the spectra: lower noise floor, wider resonances and smaller shaded area. Because the shaded area corresponds to the mean-square membrane displacement, it directly measures the effective temperature of the mode. At a drive intensity with 4,000



**Figure 3 | Sideband cooling the mechanical mode to the ground state.** **a**, The displacement noise spectra and Lorentzian fits (shaded regions) for five different drive powers. With higher power, the mechanical mode is both damped (larger linewidth) and cooled (smaller area) by the radiation pressure forces. **b**, Over a broader frequency span, the normalized sideband noise spectra clearly show both the narrow mechanical peak and a broader cavity peak due to finite occupancy of the mechanical and electrical modes, respectively. A small, but resolvable, thermal population of the cavity appears as the drive power increases, setting the limit for the final occupancy of the coupled

optomechanical system. At the highest drive power, the coupling rate between the mechanical oscillator and the microwave cavity exceeds the intrinsic dissipation of either mode, and the system hybridizes into optomechanical normal modes. **c**, Starting in thermal equilibrium with the cryostat at  $T = 20$  mK, sideband cooling reduces the thermal occupancy of the mechanical mode from  $n_m = 40$  into the quantum regime, reaching a minimum of  $n_m = 0.34 \pm 0.05$ . Error bars, s.d. These data demonstrate that the parametric interaction between photons and phonons can initialize the strongly coupled, electromechanical system in its quantum ground state.

photons in the cavity, the thermal occupation is reduced below one quantum of mechanical motion, entering the quantum regime.

Observing the noise spectrum over a broader frequency range reveals that in addition to the mechanical Lorentzian peak with linewidth  $\Gamma'_m$ , there is also a Lorentzian peak with linewidth  $\kappa$  whose area corresponds to the finite thermal occupation  $n_c$  of the cavity. Over this range, it is no longer valid to evaluate the cavity parameters at a single frequency to infer the spectrum in units of  $S_x$ . Instead, Fig. 3b shows the noise spectrum in units of sideband power normalized by the power at the drive frequency,  $S/P_o$ . These two sources of noise, originating from either the mechanical or the electrical mode, interfere with each other and result in noise squashing<sup>14</sup> and eventually normal-mode splitting<sup>28</sup> once  $2g > \kappa/\sqrt{2}$ . Using a quantum-mechanical description applied to our circuit<sup>14,24</sup>, the expected noise spectrum as a function of frequency  $\omega$  is:

$$S = \hbar\omega \left( \frac{1}{2} + n_{\text{add}} + 2\kappa_{\text{ex}} \frac{\kappa n_c (\Gamma_m^2 + 4\delta^2) + 4\Gamma_m n_m^T g^2}{|4g^2 + (\kappa + 2j(\delta + \tilde{\Delta}))(\Gamma_m + 2j\delta)|^2} \right) \quad (1)$$

where  $j = \sqrt{-1}$ ,  $\delta = \omega - \Omega_m$ ,  $\tilde{\Delta} = \omega_d + \Omega_m - \omega_c$ , and  $n_{\text{add}}$  is added noise of the microwave measurement expressed as an equivalent number of microwave photons. Figure 3b shows the measured spectra and corresponding fits (shaded regions) to equation (1) as the electromechanical system evolves first into the quantum regime ( $n_m, n_c < 1$ ) and then into the strong-coupling regime ( $2g > \kappa/2$ ). The results are summarized in Fig. 3c, where the thermal occupancy of both the mechanical and electrical modes is shown as a function of  $n_d$ . For low drive power, the cavity shows no resolvable thermal population (to within our measurement uncertainty of 0.05 quanta), as expected for a 7.5-GHz mode at 20 mK. Although it is unclear whether the observed population at higher drive power is a consequence of direct heating of the substrate, heating of the microwave attenuators preceding the circuit, or intrinsic cavity frequency noise, we have determined that it is not the result of frequency or amplitude noise of our microwave generator, as this noise is reduced far below the microwave shot-noise level with narrow-band filtering and cryogenic attenuation (see Supplementary Information). Sideband cooling can never reduce the occupancy of the mechanical mode below that of the cavity. Therefore, in order for the system to access the quantum regime, the thermal population of the cavity must remain less than one quantum. Assuming  $\Omega_m \gg \kappa$ , the final occupancy of a mechanical mode is<sup>28</sup>:

$$n_m = n_m^T \left( \frac{\Gamma_m}{\kappa} \frac{4g^2 + \kappa^2}{4g^2 + \kappa\Gamma_m} \right) + n_c \left( \frac{4g^2}{4g^2 + \kappa\Gamma_m} \right) \quad (2)$$

This equation shows that for moderate coupling ( $\sqrt{\kappa\Gamma_m} \ll g \ll \kappa$ ) the cooling of the mechanical mode is linear in the number of drive photons. Beyond this regime, the onset of normal-mode splitting abates further cooling. Here the mechanical cooling rate becomes limited not by the coupling between the mechanical mode and the cavity, but instead by the coupling rate  $\kappa$  between the cavity and its environment<sup>28</sup>. Thus, the final occupancy of the mechanical mode can never be reduced to lower than  $n_m^T \Gamma_m / \kappa$ , and a stronger parametric drive will only increase the Rabi frequency at which the thermal excitations oscillate between the cavity and mechanical modes. For our device, as the coupling is increased, we first cool to the ground state and then enter the strong-coupling regime ( $n_m^T \Gamma_m < \kappa < g$ ). Once  $n_d$  exceeds  $2 \times 10^4$ , the mechanical occupancy converges towards the cavity population, reaching a minimum of  $0.34 \pm 0.05$  quanta. At the highest power drive ( $n_d = 2 \times 10^5$ ), the mechanical mode has hybridized with the cavity, resulting in the normal-mode splitting characteristic of the strong-coupling regime<sup>16,29</sup>. This level of coupling is required to use the hybrid system for quantum information processing. The strong-coupling regime ensures that quantum states of this combined system may be manipulated faster than they decohere from spurious interaction with either electromagnetic or mechanical environments.

Taken together, the measurements shown in Figs 2 and 3 quantify the overall measurement efficiency of the system. The Heisenberg limit

requires that a continuous displacement measurement is necessarily accompanied by a back-action force<sup>3,13,24</sup>, such that  $\sqrt{S_x^{\text{imp}} S_F^{\text{ba}}} \geq \hbar$ , where  $S_F^{\text{ba}}$  is the force noise spectral density from back-action alone. From the thermal occupancy and damping rate of the mechanical mode, we extract a total force spectral density  $S_F^{\text{tot}} = 4\hbar\Omega_m m \Gamma'_m (n_m + 1/2)$ . By attributing all of  $S_F^{\text{tot}}$  to back-action, we can place a conservative upper bound on the imprecision-back-action product with  $\sqrt{S_x^{\text{imp}} S_F^{\text{tot}}} / \hbar = \sqrt{n_{\text{imp}} (n_m + 1/2)} \leq 5.1 \pm 0.4$ . The minimum achievable value with our detuned probe tone (see Supplementary Information) is  $\sqrt{S_x^{\text{imp}} S_F^{\text{ba}}} = \hbar\sqrt{2}$ , making this experiment a factor of 3.6 away from the Heisenberg-limit for displacement detection, the narrowest gap achieved to date<sup>15,24</sup>.

Looking forward, this technology offers a feasible route to achieve many of the long-standing goals for mechanical quantum systems. Whereas the resolved-sideband regime is well-suited for efficient sideband cooling, it makes a simultaneous measurement of both the upper and lower sideband difficult. By simply increasing the bandwidth of the cavity, future experiments could feasibly measure the zero-point motion of the mechanical mode, as well as observe the fundamental asymmetry between the rates of emission and absorption of phonons<sup>1</sup>. Other prospects include quantum non-demolition measurements<sup>3</sup> and the generation of entangled states of mechanical motion<sup>17,18</sup>. Furthermore, combining this device with a superconducting qubit<sup>21</sup> would allow for the direct measurement of the mechanical oscillator's energy states and the preparation of arbitrary quantum states of mechanical motion<sup>30</sup>. Because the interaction between the 10.6-MHz mechanical mode and the 7.5-GHz microwave cavity is parametric, the coupling strength is inherently tunable, and can be turned on and off quickly. Thus, once a quantum state is transferred into the mechanical mode, it can be stored there for a time  $\tau_{\text{th}} = 1/(n_m^T \Gamma_m) > 100 \mu\text{s}$  before absorbing one thermal phonon from its environment. As this timescale is much longer than typical coherence times of superconducting qubits, mechanical modes offer the potential for delay and storage of quantum information. Lastly, mechanical objects provide a generic system for interacting with a wide range of different physical systems—ranging from magnetic spins to optical photons—leading towards future methods for engineering the coherent transfer of quantum information between vastly different forms of quanta<sup>20</sup>.

The power and versatility of sideband cooling techniques have now been used to bring a high quality, macroscopic ( $\sim 10^{12}$  atoms) low-frequency mechanical oscillator into the quantum regime. This electromechanical system simultaneously demonstrates ground-state preparation, strong-coupling and near quantum-limited position detection, paving the way to accessing the quantum nature of long-lived motional states.

Received 22 February; accepted 2 June 2011.

Published online 6 July 2011.

1. Diedrich, F., Bergquist, J. C., Itano, W. M. & Wineland, D. J. Laser cooling to the zero-point energy of motion. *Phys. Rev. Lett.* **62**, 403–406 (1989).
2. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
3. Braginsky, V. B. & Khalili, F. Y. *Quantum Measurement* (Cambridge Univ. Press, 1992).
4. Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: back-action at the mesoscale. *Science* **321**, 1172–1176 (2008).
5. Braginsky, V. B., Manukin, A. B. & Tikhonov, M. Y. Investigation of dissipative ponderomotive effects of electromagnetic radiation. *Sov. Phys. JETP* **31**, 829–830 (1970).
6. Blair, D. G. *et al.* High sensitivity gravitational wave antenna with parametric transducer readout. *Phys. Rev. Lett.* **74**, 1908–1911 (1995).
7. Cuthbertson, B. D., Tobar, M. E., Ivanov, E. N. & Blair, D. G. Parametric back-action effects in a high-Q cryogenic sapphire transducer. *Rev. Sci. Instrum.* **67**, 2435–2442 (1996).
8. Teufel, J. D., Harlow, J. W., Regal, C. A. & Lehnert, K. W. Dynamical backaction of microwave fields on a nanomechanical oscillator. *Phys. Rev. Lett.* **101**, 197203 (2008).



9. Thompson, J. D. *et al.* Strong dispersive coupling of a high-finesse cavity to a micromechanical membrane. *Nature* **452**, 72–75 (2008).
10. Gröblacher, S. *et al.* Demonstration of an ultracold micro-optomechanical oscillator in a cryogenic cavity. *Nature Phys.* **5**, 485–488 (2009).
11. Park, Y.-S. & Wang, H. Resolved-sideband and cryogenic cooling of an optomechanical resonator. *Nature Phys.* **5**, 489–493 (2009).
12. Lin, Q., Rosenberg, J., Jiang, X., Vahala, K. J. & Painter, O. Mechanical oscillation and cooling actuated by the optical gradient force. *Phys. Rev. Lett.* **103**, 103601 (2009).
13. Schliesser, A., Arcizet, O., Rivière, R., Anetsberger, G. & Kippenberg, T. J. Resolved-sideband cooling and position measurement of a micromechanical oscillator close to the Heisenberg uncertainty limit. *Nature Phys.* **5**, 509–514 (2009).
14. Rocheleau, T. *et al.* Preparation and detection of a mechanical resonator near the ground state of motion. *Nature* **463**, 72–75 (2010).
15. Rivière, R. *et al.* Optomechanical sideband cooling of a micromechanical oscillator close to the quantum ground state. Preprint at (<http://arXiv.org/abs/1011.0290>) (2010).
16. Teufel, J. D. *et al.* Circuit cavity electromechanics in the strong coupling regime. *Nature* **471**, 204–208 (2011).
17. Bose, S., Jacobs, K. & Knight, P. L. Preparation of nonclassical states in cavities with a moving mirror. *Phys. Rev. A* **56**, 4175–4186 (1997).
18. Mancini, S., Man'ko, V. I. & Tombesi, P. Ponderomotive control of quantum macroscopic coherence. *Phys. Rev. A* **55**, 3042–3050 (1997).
19. Marshall, W., Simon, C., Penrose, R. & Bouwmeester, D. Towards quantum superpositions of a mirror. *Phys. Rev. Lett.* **91**, 130401 (2003).
20. Regal, C. A. & Lehnert, K. W. From cavity electromechanics to cavity optomechanics. *J. Phys. Conf. Ser.* **264**, 012025 (2011).
21. O'Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
22. Marquardt, F., Chen, J. P., Clerk, A. A. & Girvin, S. M. Quantum theory of cavity-assisted sideband cooling of mechanical motion. *Phys. Rev. Lett.* **99**, 093902 (2007).
23. Wilson-Rae, I., Nooshi, N., Zwerger, W. & Kippenberg, T. J. Theory of ground state cooling of a mechanical oscillator using dynamical backaction. *Phys. Rev. Lett.* **99**, 093901 (2007).
24. Clerk, A. A., Devoret, M. H., Girvin, S. M., Marquardt, F. & Schoelkopf, R. J. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.* **82**, 1155–1208 (2010).
25. Cicak, K. *et al.* Low-loss superconducting resonant circuits using vacuum-gap-based microwave components. *Appl. Phys. Lett.* **96**, 093502 (2010).
26. Castellanos-Beltran, M. A., Irwin, K. D., Hilton, G. C., Vale, L. R. & Lehnert, K. W. Amplification and squeezing of quantum noise with a tunable Josephson metamaterial. *Nature Phys.* **4**, 929–931 (2008).
27. Teufel, J. D., Donner, T., Castellanos-Beltran, M. A., Harlow, J. W. & Lehnert, K. W. Nanomechanical motion measured with an imprecision below that at the standard quantum limit. *Nature Nanotechnol.* **4**, 820–823 (2009).
28. Dobrindt, J. M., Wilson-Rae, I. & Kippenberg, T. J. Parametric normal-mode splitting in cavity optomechanics. *Phys. Rev. Lett.* **101**, 263602 (2008).
29. Gröblacher, S., Hammerer, K., Vanner, M. R. & Aspelmeyer, M. Observation of strong coupling between a micromechanical resonator and an optical cavity field. *Nature* **460**, 724–727 (2009).
30. Hofheinz, M. *et al.* Synthesizing arbitrary quantum states in a superconducting resonator. *Nature* **459**, 546–549 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. W. Sanders for taking the micrograph in Fig. 1 and the JILA instrument shop for fabrication and design of the cavity filter. This work was supported by NIST and the DARPA QuASAR programme. T.D. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG). This Letter is a contribution of the US government and not subject to copyright.

**Author Contributions** J.D.T. and R.W.S. conceived the device. J.D.T. designed the circuit. J.D.T. and D.L. fabricated the devices. J.D.T. and T.D. set up the experiment, performed the measurements and analysed the data. J.D.T., T.D., R.W.S. and K.W.L. discussed the results and wrote the manuscript. All authors provided experimental support and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.D.T. ([john.teufel@nist.gov](mailto:john.teufel@nist.gov)).

# Inkjet printing of single-crystal films

Hiromi Minemawari<sup>1</sup>, Toshikazu Yamada<sup>1</sup>, Hiroyuki Matsui<sup>1</sup>, Jun'ya Tsutsumi<sup>1</sup>, Simon Haas<sup>1</sup>, Ryosuke Chiba<sup>1,2</sup>, Reiji Kumai<sup>1,3</sup> & Tatsuo Hasegawa<sup>1</sup>

The use of single crystals has been fundamental to the development of semiconductor microelectronics and solid-state science<sup>1</sup>. Whether based on inorganic<sup>2–5</sup> or organic<sup>6–8</sup> materials, the devices that show the highest performance rely on single-crystal interfaces, with their nearly perfect translational symmetry and exceptionally high chemical purity. Attention has recently been focused on developing simple ways of producing electronic devices by means of printing technologies. 'Printed electronics' is being explored for the manufacture of large-area and flexible electronic devices by the patterned application of functional inks containing soluble or dispersed semiconducting materials<sup>9–11</sup>. However, because of the strong self-organizing tendency of the deposited materials<sup>12,13</sup>, the production of semiconducting thin films of high crystallinity (indispensable for realizing high carrier mobility) may be incompatible with conventional printing processes. Here we develop a method that combines the technique of antisolvent crystallization<sup>14</sup> with inkjet printing to produce organic semiconducting thin films of high crystallinity. Specifically, we show that mixing fine droplets of an antisolvent and a solution of an active semiconducting component within a confined area on an amorphous substrate can trigger the controlled formation of exceptionally uniform single-crystal or polycrystalline thin films that grow at the liquid–air interfaces. Using this approach, we have printed single crystals of the organic semiconductor 2,7-dioctyl[1]benzothieno[3,2-*b*] [1]benzothiophene (C<sub>8</sub>-BTBT) (ref. 15), yielding thin-film transistors with average carrier mobilities as high as 16.4 cm<sup>2</sup> V<sup>−1</sup> s<sup>−1</sup>. This printing technique constitutes a major step towards the use of high-performance single-crystal semiconductor devices for large-area and flexible electronics applications.

Antisolvent crystallization is recognized as the best method of achieving controlled and scalable solidification, which is useful in pharmaceutical science<sup>14</sup>, for example. To achieve this, an 'antisolvent' (a liquid in which a substance is insoluble) is added to the solution of the substance in a solvent that is miscible with the antisolvent. Here we make use of this concept in microliquid inkjet printing processes.

A solution of a semiconductor and an antisolvent for the semiconductor are used as the two kinds of ink; the inks are individually printed at arbitrary positions to form a microliquid intermixture between the inks on the top of substrates. We found that optimized printing conditions enable controlled formation of patterned single-crystal thin films having molecularly flat surfaces, in contrast to conventional inkjet printing processes that produce films with a non-uniform thickness distribution. This is a conceptual extension of the 'double-shot' inkjet printing process that was developed to produce films of charge-transfer compounds that are hardly soluble<sup>16,17</sup>. We used 1,2-dichlorobenzene (DCB) as the solvent and *N,N*-dimethylformamide (DMF) as the antisolvent for the semiconductor C<sub>8</sub>-BTBT. These organic liquids show very different solubilities for C<sub>8</sub>-BTBT (the solubility at 20 °C is 400 times higher in DCB than in DMF), but have similar boiling points and are miscible with one another.

A schematic representation of this printing process is shown in Fig. 1a. We used silicon wafers with 100-nm-thick silicon dioxide layers

as substrates. We produced the wetting/non-wetting surface patterning on the silicon dioxide layers by using a combination of ultraviolet/ozone treatment, hexamethyldisilazane treatment, and photoresist patterning<sup>18</sup>. We used a piezoelectric inkjet printing apparatus with double inkjet printing heads, from which a droplet of 60 picolitres is ejected at a repetition frequency of 500 Hz. In the process, the antisolvent ink (pure anhydrous DMF) is printed first and then overprinted with the solution ink (a 28 mM solution of C<sub>8</sub>-BTBT in DCB). In the formation of all the pieces of film shown in Fig. 1b, 42 shots of antisolvent ink were printed first and then 6 shots of solution ink were overprinted, all within a second. The deposited droplets are confined and intermixed in a predefined hydrophilic area on the upper surface of the substrate.

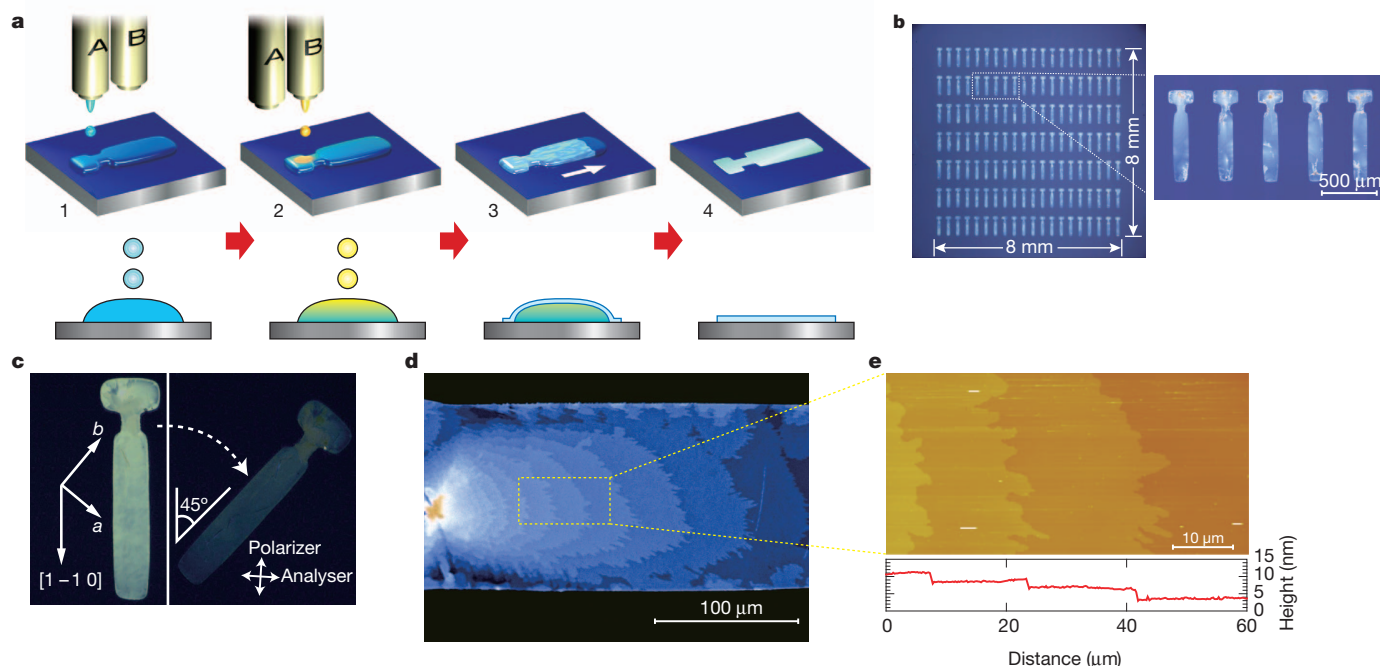
During the initial stages of film formation, tiny floating bodies begin to form at the surface of the liquid and can be seen in microscope images (Supplementary Movie). Each floating body acts as a nucleus for further crystallization and undergoes subsequent growth to form a larger floating body. These bodies eventually cover the entire surface of the droplet (step 3 in Fig. 1a). A few creases can be seen on the surfaces of the droplets during liquid evaporation, indicating the solid nature of the films (step 4 and Supplementary Fig. 1)<sup>19</sup>.

Although nuclei are generated randomly, mostly at the perimeters of the deposited droplet (solid–liquid–air interfaces), we found that nucleation can be controlled through appropriate design of the droplet configuration, which is shaped by the predefined hydrophilic area as well as by the ink volume. For example, a hydrophilic area containing a protuberance, as presented in Fig. 1b, was quite effective in causing local seeding of floating bodies in the protrusive area. We propose that local seeding is associated with the comparatively higher rate of solvent evaporation in areas with a high surface area-to-volume ratio. After seeding, the growing front moves slowly to the other end of the droplet until the large single-domain floating body covers the entire liquid–air interface (see Supplementary Movie).

The solvent then evaporates very slowly, taking about 10–50 times longer than is the case without the solute, most probably because the droplet is completely covered by the solid film. During this slow evaporation, the creases in the films become smoothed out, and films with thickness of about 30–200 nm are eventually obtained on the amorphous substrate. The film adheres tightly to the substrate. The morphology of the films as well as their single-domain nature depends on a variety of printing conditions, such as substrate temperature, the concentration and volume of the solution, the solution–antisolvent ratio and the shape of the hydrophilic area on which the droplet is deposited.

The thickness profile of the film differs markedly from that of conventional inkjet printing deposits. Conventional inkjet printing is known to produce a characteristic thickness distribution in which both ends of the deposit are considerably thicker than its centre, known as the 'coffee-ring effect' (see Supplementary Fig. 2)<sup>20</sup>. The uniform nature of the deposits produced by our process can be ascribed to temporal discrimination between solute crystallization and solvent evaporation within the deposited droplet (see Supplementary Fig. 1)<sup>16</sup>. The occurrence of supersaturation in the intermixed microliquid droplet results

<sup>1</sup>National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central 4, Tsukuba 305-8562, Japan. <sup>2</sup>Department of Applied Physics, The University of Tokyo, Hongo 113-8656, Japan. <sup>3</sup>CMRC, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba 305-0801, Japan.



**Figure 1 | Inkjet printing of organic single-crystal thin films.** **a**, Schematic of the process. Antisolvent ink (A) is first inkjet-printed (step 1), and then solution ink (B) is overprinted sequentially to form intermixed droplets confined to a predefined area (step 2). Semiconducting thin films grow at liquid–air interfaces of the droplet (step 3), before the solvent fully evaporates (step 4).

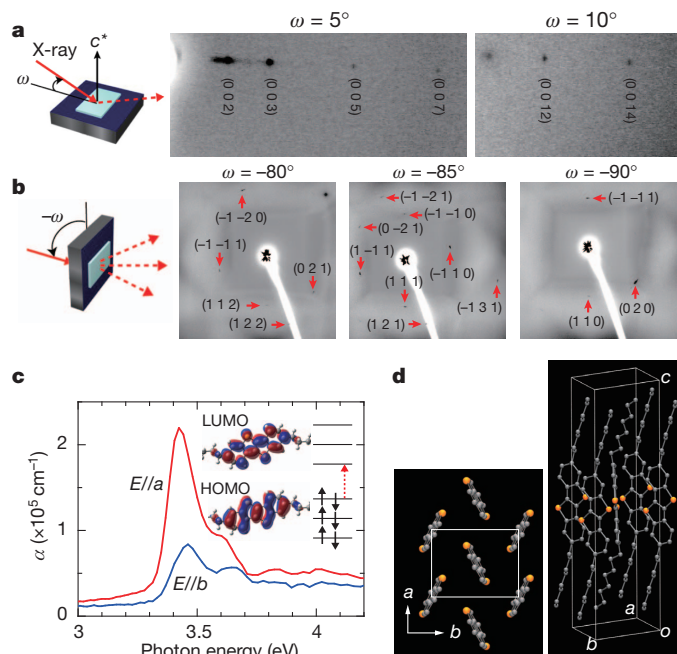
in solute crystallization before solvent evaporation. In the microscope images of the films shown in Fig. 1d, we can see stripe-like features with intervals of several micrometres to several tens of micrometres. Atomic-force microscopy showed that the stripes are associated with the height of the molecular step, which is estimated to be about 2.6–2.8 nm (Fig. 1e)<sup>21</sup>. This value is consistent with the thickness of one molecular layer of C<sub>8</sub>-BTBT ( $c\sin\beta = 2.92$  nm, where  $c$  and  $\beta$  are the unit cell parameters) (ref. 22). We conclude that the stripe-like features are associated with the step-and-terrace structure of C<sub>8</sub>-BTBT.

In images recorded through crossed Nicol prisms, the colour of almost the entire film changes from bright to dark, simultaneously, on rotating the film about an axis perpendicular to the substrate (Fig. 1c). In addition, when we use hydrophilic areas with different configurations such as a simple square, rectangle or circle, we obtained polycrystalline films composed of some crystal domains (see Supplementary Fig. 3). From these observations, we conclude that with appropriate design of both the droplet shape and printing conditions, single-domain crystal films that cover nearly the whole region of the printed deposits could be produced with high probability (Supplementary Fig. 4). We also noticed that the step-and-terrace structures in Fig. 1d form concentric ellipses, and propose that this feature is formed by epitaxial growth on top of thinner single-domain crystal films at a later stage (see Supplementary Fig. 5).

X-ray diffraction data for the films are shown in Fig. 2a and b. The observed out-of-plane diffraction spots are consistent with a molecular layer structure that is parallel to the  $a$  and  $b$  axes. The observation of Bragg reflections up to 14th order indicates that the films have a highly crystalline nature. At high incident angles of the X-rays, we observed 16 diffraction spots that could be ascribed to Bragg reflections with indices that include an in-plane component (Fig. 2b), where the refined unit cell—monoclinic  $P2_1/a$ ,  $a = 5.91(15)$  Å,  $b = 7.88(1)$  Å,  $c = 29.12(19)$  Å,  $\beta = 91.0(8)^\circ$ ,  $V = 1357(4)$  Å<sup>3</sup>—is consistent with that of the bulk crystal<sup>22</sup>. These results provide unambiguous evidence that the films are single-crystalline with a long-range translational symmetry.

The data show that the growth direction is parallel to  $[1 -1 0]$  in many (about 60–70%) of the deposited films (Fig. 1c). On the other

**b**, Micrographs of a  $20 \times 7$  array of inkjet-printed C<sub>8</sub>-BTBT single-crystal thin films. **c**, Crossed Nicols polarized micrographs of the film. **d**, Expanded micrograph of the film, showing stripes caused by molecular-layer steps. **e**, Atomic-force microscopy image and the height profile (below) showing the step-and-terrace structure on the film surfaces.



**Figure 2 | Synchrotron-radiated single-crystal X-ray diffraction and polarized absorption spectra.** Oscillation photographs for out-of-plane diffraction (**a**) and for high-incident-angle diffraction (**b**) of inkjet-printed C<sub>8</sub>-BTBT single-crystal thin films, where  $\omega$  is the incident angle. The Bragg reflections observed in **b** correspond to the indices, which contain in-plane components. The refined unit cell obtained from the reflections is consistent with that of the bulk crystal. **c**, Polarized optical absorption spectra with coefficient  $\alpha$  and with polarization parallel to the  $a$  and  $b$  axes in the single-crystal film, demonstrating optical anisotropy with regard to these principal axes. **d**, View of the molecular arrangement of C<sub>8</sub>-BTBT in the crystal<sup>21</sup>.



hand, the bright-to-dark images observed through the crossed Nicol prisms originate from in-plane optical anisotropy of the single-crystal films. Figure 2c shows the polarized optical absorption spectra of our single-crystal films with the electric field of the light parallel to the *a* or the *b* axis. The spectra show a clear anisotropy in their absorption intensity and peak energy. The absorption intensity is much higher along the *a* axis and peaks at 3.43 eV, whereas the absorption intensity along the *b* axis is comparatively weak and peaks at a higher energy of 3.47 eV. We note that the transition dipole for the lowest electronic excitation between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is polarized parallel to the molecular plane (Fig. 2c). The difference in the absorption intensity can be clearly ascribed to the orientation of the molecular planes within the *a*–*b* plane. In contrast, the difference in peak energies is due to Davydov splitting along the *a* and *b* axes; this is characteristic of herringbone-type molecular arrangements within single-crystal films, as observed in anthracene<sup>23</sup> or pentacene<sup>24</sup>.

Field-effect devices were fabricated for the single-crystal films with a top-contact/top-gate geometry, composed of 30-nm Au films as the source and drain electrodes, and films of parylene C (capacitance per unit area of  $C = 4.2 \text{ nF cm}^{-2}$ ) as the gate dielectric layers. The typical channel width and length were 145  $\mu\text{m}$  and 100  $\mu\text{m}$ , respectively. The direct-current field-effect characteristics at room temperature (300 K) were measured in an argon-filled glove box. The transfer and output characteristics of this device are shown in Fig. 3. The mobility in the saturation regime reaches  $16.4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$  on average, and the maximum value is as high as  $31.3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . The on/off current ratio is  $10^5$ – $10^7$ , and the subthreshold slope was about 2 V per decade with a threshold voltage of about  $-10 \text{ V}$ . Injection barriers at the source/drain contacts may have remained, as manifested by the slightly nonlinear source/drain current–voltage ( $I_{\text{sd}} - V_{\text{sd}}$ ) dependence at low voltages. Hardly any current hysteresis was observed in the transfer and output characteristics, where the shift in the threshold voltage from forward to reverse sweeps was less than 0.1 V. This

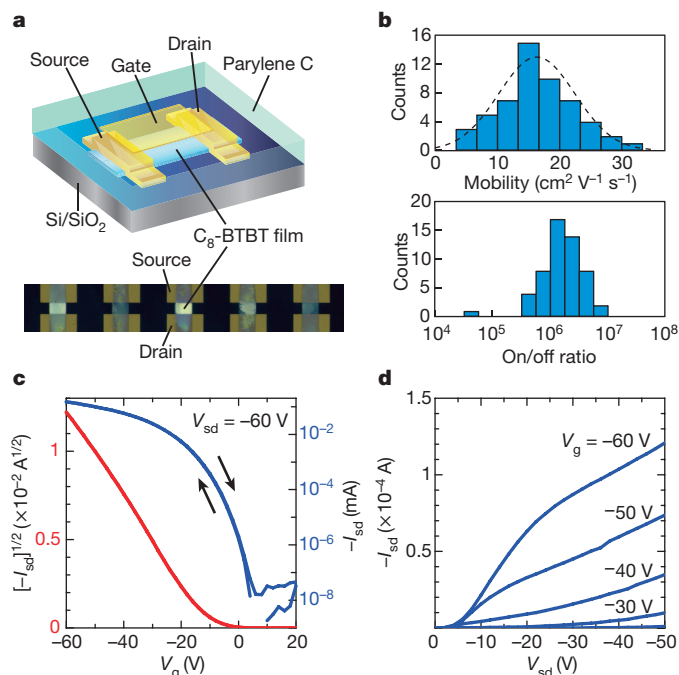
feature is probably associated with the negligible charge-trapping effects between the single-crystal surface of  $\text{C}_8$ -BTBT and the parylene gate dielectric layer. The slope of the transfer curve (Fig. 3c) presents a distinct kink feature, as reported in other organic single-crystal devices<sup>8</sup>, which clearly demonstrates the high quality of the semiconductor–insulator interface. We also found that the characteristics were not influenced by the existence of a few domain boundaries and were not degraded by more than 10% after the films were kept in air for 8 months.

This device performance is much higher than the previous report for  $\text{C}_8$ -BTBT<sup>21</sup> and is comparable to the highest performance obtained (for a rubrene single-crystal device<sup>8</sup>). We consider that the following characteristics of the film formation process are important to achieve high-quality single-crystal film: (1) the liquid–air interfaces need to be ideal locations for diffusion and self-organization of organic molecules (as for Langmuir–Blodgett films<sup>25</sup>) and (2) the gradual growth of single-crystal films is only possible because of the fluidic nature of the microliquid droplet in which laminar flow dominates over turbulent flow<sup>26</sup>. The technique should be applicable to a broad class of functional soluble materials.

The rather broad distribution of apparent mobility (Fig. 3b) indicates that further improvements of our technique should be possible, in areas such as ink composition, the optimization of equipment and the environment, and also subsequent device processing. For example, there is plenty of scope for improving the source/drain contacts. Nonetheless, we believe that this drop-on-demand, non-vacuum and room-temperature printing process of patterned single-crystal semiconductor films is in principle a useful new way of producing transistor arrays on top of plastic substrates, which is indispensable for realizing large-area, light-weight and high-speed electronic products.

Received 20 February; accepted 14 June 2011.

Published online 13 July 2011.



**Figure 3 | Transistor characteristics for the inkjet-printed  $\text{C}_8$ -BTBT single-crystal thin films.** **a**, Schematic of the device structure and micrograph of the thin-film transistors. **b**, Distribution of mobility and on/off ratio measured over 54 transistors. Average mobility is  $16.4 \pm 6.1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ . **c**, Transfer characteristics at  $V_{\text{sd}} = -60 \text{ V}$ . **d**, Output characteristics at various gate voltages  $V_g$ .

- Teal, G. K. Single crystals of germanium and silicon—basic to the transistor and integrated circuit. *IEEE Trans. Electron. Dev.* **23**, 621–639 (1976).
- Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
- De Poortere, E. P. *et al.* Enhanced electron mobility and high order fractional quantum Hall states in AlAs quantum wells. *Appl. Phys. Lett.* **80**, 1583–1585 (2002).
- Ohtomo, A. & Hwang, H. Y. A high-mobility electron gas at the  $\text{LaAlO}_3/\text{SrTiO}_3$  heterointerface. *Nature* **427**, 423–426 (2004).
- Tsukazaki, A. *et al.* Observation of the fractional quantum Hall effect in an oxide. *Nature Mater.* **9**, 889–893 (2010).
- Sundar, V. C. *et al.* Elastomeric transistor stamps: reversible probing of charge transport in organic crystals. *Science* **303**, 1644–1646 (2004).
- Briseno, A. L. *et al.* Patterning organic single-crystal transistor arrays. *Nature* **444**, 913–917 (2006).
- Takeya, J. *et al.* Very high-mobility organic single-crystal transistors with in-crystal conduction channel. *Appl. Phys. Lett.* **90**, 102120 (2007).
- Yan, H. *et al.* A high-mobility electron-transporting polymer for printed transistors. *Nature* **457**, 679–686 (2009).
- Rivnay, J. *et al.* Large modulation of carrier transport by grain-boundary molecular packing and microstructure in organic thin films. *Nature Mater.* **8**, 952–958 (2009).
- Noh, Y.-Y., Zhao, N., Caironi, M. & Sirringhaus, H. Downscaling of self-aligned, all-printed polymer thin-film transistors. *Nature Nanotechnol.* **2**, 784–789 (2007).
- Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295**, 2418–2421 (2002).
- Lim, J. A., Lee, H. S., Lee, W. H. & Cho, K. Control of the morphology and structural development of solution-processed functionalized acenes for high-performance organic transistors. *Adv. Funct. Mater.* **19**, 1515–1525 (2009).
- Tung, H.-H., Paul, E. L., Midler, M. & McCauley, J. A. *Crystallization of Organic Compounds: An Industrial Perspective* 179–205 (Wiley, 2009).
- Ebata, H. *et al.* Highly soluble [1]benzothieno[3,2-*b*]benzothiophene (BTBT) derivatives for high-performance, solution-processed organic field-effect transistors. *J. Am. Chem. Soc.* **129**, 15732–15733 (2007).
- Hiraoka, M. *et al.* On-substrate synthesis of molecular conductor films and circuits. *Adv. Mater.* **19**, 3248–3251 (2007).
- Hasegawa, T., Hiraoka, M. & Yamada, T. Double-shot inkjet printing of donor–acceptor-type organic charge-transfer complexes: wet/nonwet definition and its use for contact engineering. *Thin Solid Films* **518**, 3988–3991 (2010).
- Sirringhaus, H. *et al.* High-resolution inkjet printing of all-polymer transistor circuits. *Science* **290**, 2123–2126 (2000).
- Chen, Z., Liu, M., Liu, G.-y. & Tan, L. Evaporation induced two-dimensional buckling within liquid droplet. *Appl. Phys. Lett.* **95**, 223104 (2009).

20. Deegan, R. D. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).
21. Uemura, T., Hirose, Y., Uno, M., Takimiya, K. & Takeya, J. Very high mobility in solution-processed organic thin-film transistors of highly ordered [1]benzothieno[3,2-*b*]benzothiophene derivatives. *Appl. Phys. Expr.* **2**, 111501 (2009).
22. Izawa, T., Miyazaki, E. & Takimiya, K. Molecular ordering of high-performance soluble molecular semiconductors and re-evaluation of their field-effect transistor characteristics. *Adv. Mater.* **20**, 3388–3392 (2008).
23. Pope, M. & Swenberg, C. E. *Electronic Processes in Organic Crystals and Polymers* 2nd edn, 59–66 (Oxford Science, 1999).
24. Faltermeier, D., Gompf, B., Dressel, M., Tripathi, A. K. & Pflaum, J. Optical properties of pentacene thin films and single crystals. *Phys. Rev. B* **74**, 125416 (2006).
25. Ulman, A. *An Introduction to Ultrathin Organic Films, from Langmuir-Blodgett to Self-Assembly* 1st edn (Academic, 1991).
26. deMello, A. J. Control and detection of chemical reactions in microfluidic systems. *Nature* **442**, 394–402 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to Nippon Kayaku for providing C<sub>8</sub>-BTBT. We thank K. Takimiya and S. Horiuchi for discussions, H. Okamoto and H. Matsuzaki for help with optical measurements, K. Kobayashi for help with the X-ray measurements, and T. Iwade for help with atomic-force microscopy and measurements of device

characteristics. The synchrotron X-ray study was performed with the approval of the Photon Factory Program Advisory Committee (no. 2009S2-003). This work was supported by the New Energy and Industrial Technology Development Organization (NEDO) through a Grant for Industrial Technology Research and also by the Japan Society for the Promotion of Science (JSPS) through its Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program).

**Author Contributions** H. Minemawari was responsible for ink fabrication, inkjet printing, microscopic observations, X-ray diffraction measurements, and measurements of the device characteristics of all the films. T.Y. prepared substrates with the wet/non-wet surface patterning, assisted in inkjet printing and X-ray diffraction measurements, and performed atomic-force microscopy and device characteristics measurements. H. Matsui guided sample preparation and inkjet printing, and conducted DFT molecular orbital calculations. J.T. assisted with X-ray diffraction measurements and performed optical anisotropic absorption measurements. S.H. assisted with optical anisotropic absorption measurements. R.C. assisted in the ink fabrication. R.K. assisted with X-ray diffraction measurements. T.H. conceptualized and directed the research project, and wrote most of the manuscript. All the authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.H. (t-hasegawa@aist.go.jp).

# Neural network computation with DNA strand displacement cascades

Lulu Qian<sup>1</sup>, Erik Winfree<sup>1,2,3</sup> & Jehoshua Bruck<sup>3,4</sup>

The impressive capabilities of the mammalian brain—ranging from perception, pattern recognition and memory formation to decision making and motor activity control—have inspired their re-creation in a wide range of artificial intelligence systems for applications such as face recognition, anomaly detection, medical diagnosis and robotic vehicle control<sup>1</sup>. Yet before neuron-based brains evolved, complex biomolecular circuits provided individual cells with the ‘intelligent’ behaviour required for survival<sup>2</sup>. However, the study of how molecules can ‘think’ has not produced an equal variety of computational models and applications of artificial chemical systems. Although biomolecular systems have been hypothesized to carry out neural-network-like computations *in vivo*<sup>3,2,4</sup> and the synthesis of artificial chemical analogues has been proposed theoretically<sup>5–9</sup>, experimental work<sup>10–13</sup> has so far fallen short of fully implementing even a single neuron. Here, building on the richness of DNA computing<sup>14</sup> and strand displacement circuitry<sup>15</sup>, we show how molecular systems can exhibit autonomous brain-like behaviours. Using a simple DNA gate architecture<sup>16</sup> that allows experimental scale-up of multilayer digital circuits<sup>17</sup>, we systematically transform arbitrary linear threshold circuits<sup>18</sup> (an artificial neural network model) into DNA strand displacement cascades that function as small neural networks. Our approach even allows us to implement a Hopfield associative memory<sup>19</sup> with four fully connected artificial neurons that, after training *in silico*, remembers four single-stranded DNA patterns and recalls the most similar one when presented with an incomplete pattern. Our results suggest that DNA strand displacement cascades could be used to endow autonomous chemical systems with the capability of recognizing patterns of molecular events, making decisions and responding to the environment.

The human brain is composed of  $\sim 10^{11}$  neurons, and each has a few thousand synapses. Each synapse can receive signals from other neurons, raising or lowering the electrical potential inside the neuron. When the potential reaches its threshold, the neuron will fire and a pulse will be sent through the axon to other neurons. Among the simplest mathematical models of neurons is the perceptron, also known as the linear threshold gate<sup>1,18,20</sup>. A linear threshold gate has a number of inputs,  $x_1, x_2, \dots, x_n \in \{0, 1\}$ , which can be interpreted as arriving at synapses that each have an analogue weight,  $w_1, w_2, \dots, w_n$ . The linear threshold gate turns ‘on’ only when the weighted sum of all inputs exceeds a threshold,  $\text{th}$ . The output

$$y = \begin{cases} 1 & \text{if } \sum_{i=1}^n w_i x_i \geq \text{th} \\ 0 & \text{otherwise} \end{cases}$$

can be interpreted as the firing activity on the axon. Linear threshold gates may be used to construct multilayer circuits that are complete for Boolean functions and, more importantly, are exponentially more compact than AND–OR–NOT circuits for a wide class of functions<sup>18,21,22</sup>. Recurrent linear threshold circuits have even provided insights into brain-like computations, such as content-addressable associative memories<sup>19</sup>. A remarkable feature of brains, which is also

desirable for molecular circuits, is that complex computations can be carried out by networks with just a few layers and even with unreliable components—a feature that linear threshold circuits share<sup>23</sup>.

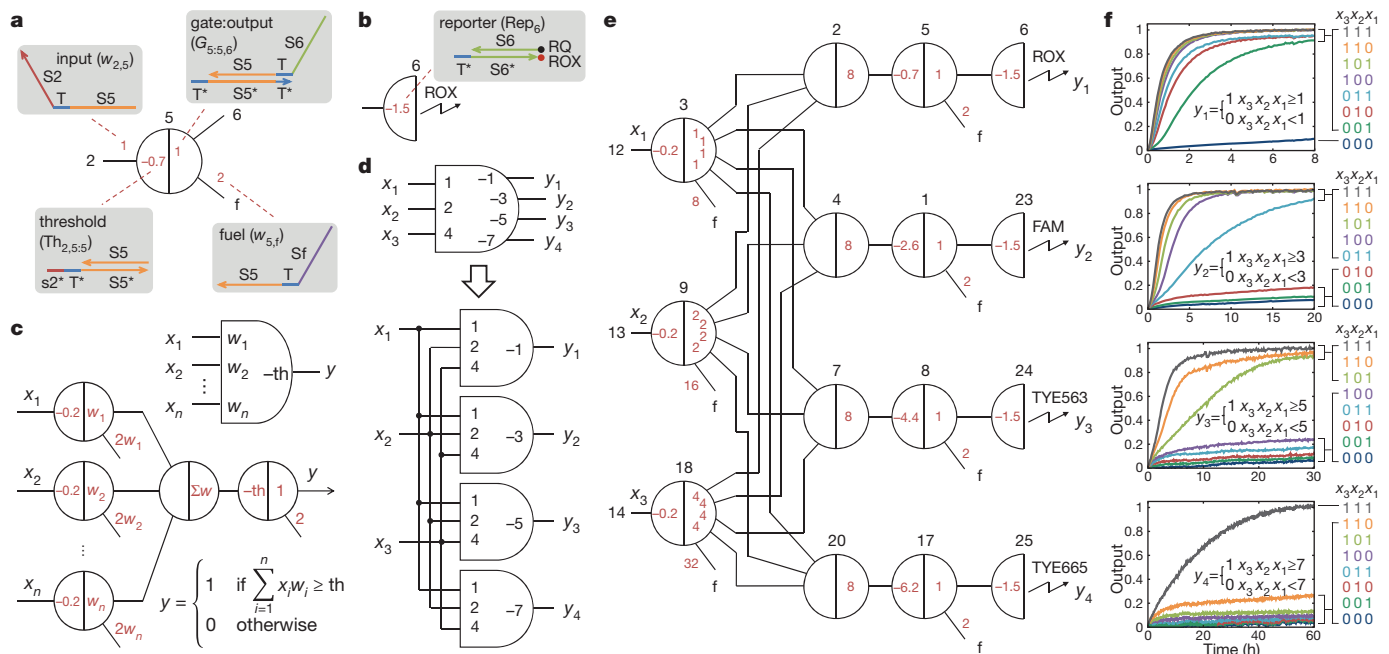
We first introduce the simple DNA gate architecture, based on what we call the ‘seesaw’ gate motif, which we use for building arbitrary linear threshold circuits. Because DNA hybridization depends primarily on the logic of Watson–Crick base-pairing, many instances of the same molecular motif can be created by assigning different sequence choices for each logical domain. The abstract diagram for the seesaw gate motif (Fig. 1a) provides a concise representation of a full DNA implementation and can be systematically translated first to the domain level, then to the sequence level, and finally to the molecular level (Supplementary Fig. 1). Each seesaw gate is a node with two sides, connected to one or more wires on each side. Each wire represents a DNA ‘signal strand’ with two long ‘recognition’ domains flanking a central short ‘toehold’ domain (for example, ‘input’ and ‘fuel’ in Fig. 1a). Each node represents a DNA gate ‘base strand’ with one central recognition domain flanked by two toehold domains. The gate base strand is always bound to a signal strand on one side or another, leaving one toehold uncovered (for example, the ‘gate:output complex’ in Fig. 1a). A DNA ‘threshold complex’ can also be associated with a node; it has a double-stranded recognition domain with an extended toehold. To read the output signal, a ‘reporter’ gate is used (Fig. 1b). The reporter is implemented as a threshold-like DNA complex with a fluorophore/quencher pair at the end of the duplex.

There are three basic reactions involved in a seesaw network (Supplementary Fig. 4). They all use the principle of toehold-mediated DNA strand displacement<sup>15</sup>, in which a single-stranded DNA binds to a partially double-stranded complex by a single-stranded toehold domain, allowing initiation of branch migration through a recognition domain with identical sequence, and ultimately resulting in replacement and release of the originally bound strand. The first reaction, seesawing, occurs when a free signal on one side of a gate releases a signal bound on the other side. A single step of seesawing results in stoichiometric exchange of equal amounts of activity from a wire on one side of a gate to a wire on the other side (for example, input releases output). Two steps of seesawing completes a catalytic cycle in which a wire on one side of a gate exchanges the activity between two wires on the other side without itself being consumed (for example, input transforms free fuel into free output, see Supplementary Fig. 5). Second, thresholding occurs when a threshold complex absorbs an impinging signal—this happens at a much faster rate than seesawing because of the extended toehold<sup>15</sup>. Third, reporting occurs when a reporter complex absorbs an impinging signal while generating a fluorescence signal.

A single linear threshold gate can be implemented using three types of seesaw gates that implement the three essential subfunctions: multiplying ( $w_i x_i$ ), integrating ( $\sum w_i x_i$ ) and thresholding ( $\sum w_i x_i \geq \text{th}$ ). (1) Multiplying gates (for example, the first layer of seesaw gates in Fig. 1c, e) have a fixed threshold of 0.2 and support multiple outputs with arbitrary weights. Each output strand has a different recognition domain on the right (5’ end) to connect to different downstream gates.

<sup>1</sup>Bioengineering, California Institute of Technology, Pasadena, California 91125, USA. <sup>2</sup>Computer Science, California Institute of Technology, Pasadena, California 91125, USA. <sup>3</sup>Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, USA. <sup>4</sup>Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA.





**Figure 1 | The seesaw gate motif and the construction of linear threshold gates.** **a**, Abstract diagram of a seesaw gate motif and its DNA implementation. Black numbers indicate the identity of each node (or the interface to that node in a larger network). Positions and signs of red numbers indicate different DNA species, while their absolute values indicate the initial relative concentrations (for details, see Supplementary Information section 1). Each species has a specific role (for example, input) within a gate and has a unique name (for example,  $w_{2,5}$ ) within a network. Coloured lines represent DNA strands at the domain level, with arrowheads marking their 3' ends and colours indicating distinct DNA subsequences. S2, S5, S6 and Sf are long recognition domains. T is a short toehold domain. T\* is the Watson–Crick complement of T, and so on.  $s2^*$  is the first few nucleotides of S2\* from the 3' end. **b**, Abstract diagram of a reporter and its DNA implementation. Fluorophore ROX is quenched by quencher RQ. **c**, A linear threshold gate and its equivalent seesaw construction. **d**, A general three-input four-output linear threshold gate. **e**, Equivalent seesaw

To clean up variations due to leaky reactions and signal decay, we require all gates to work with a digital abstraction where 'off' signals may be between 0 to 0.2, and 'on' signals may be between 0.8 to 1. If the input is 'off', all outputs will remain 0. If the input is 'on', it will exceed the threshold and catalyse the exchange of fuel and outputs. With an irreversible downstream drain, each output will continue being released until no gate:output complexes remain. Thus, each output level is set by an analogue weight—the initial amount of gate:output complex. (2) Integrating gates (for example, the second layer of seesaw gates in Fig. 1c, e) have no threshold or fuel, but support multiple inputs. All input strands have the same right recognition domain to connect to this gate, but have different left recognition domains corresponding to different upstream gates. Without fuel, the gate exhibits a stoichiometric behaviour with the output level eventually reaching the sum of all inputs. (3) Thresholding gates (for example, the third layer of seesaw gates in Fig. 1c, e) have an arbitrary threshold and an output with a fixed weight of 1. If the input exceeds the threshold, the output will turn 'on'; otherwise it will stay 'off'. To reduce circuit size in cascades, multiplying gates and thresholding gates can be combined and generalized as a fourth type, amplifying gates, that allow both an arbitrary threshold and multiple outputs with arbitrary weights (Supplementary Fig. 6). The full DNA strand displacement cascade implementing a single neuron is shown in Supplementary Fig. 7.

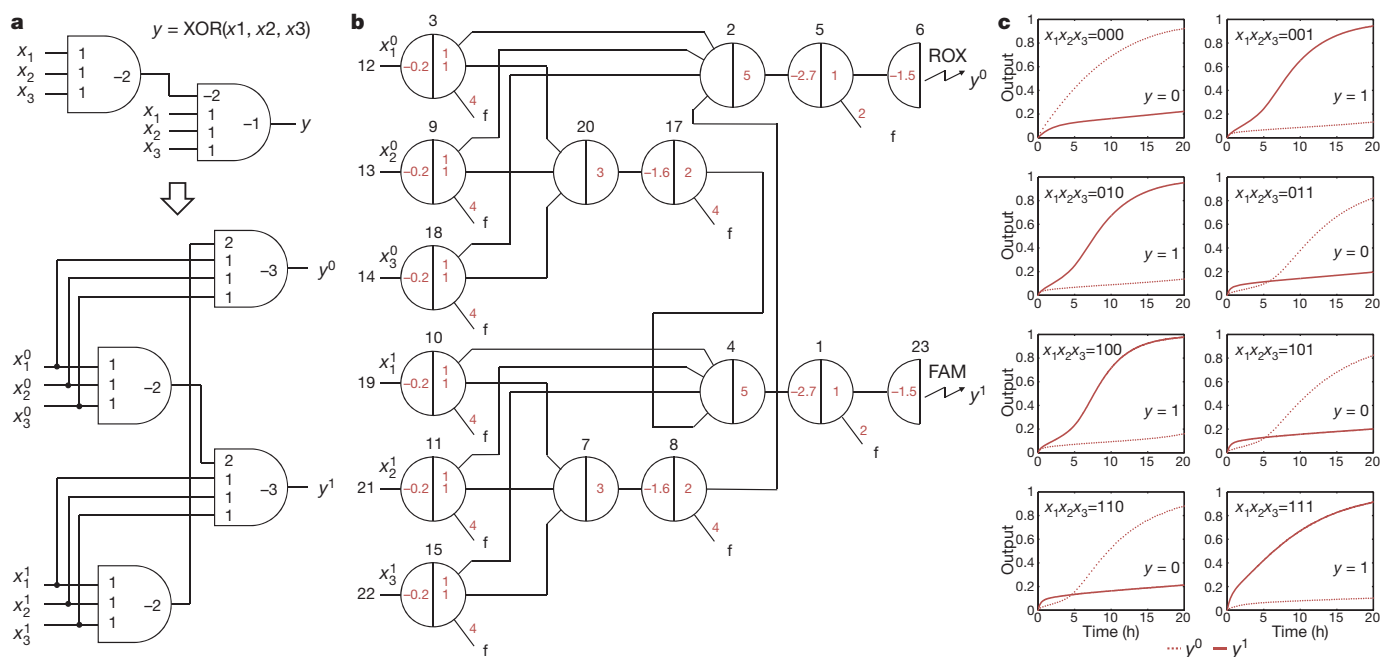
To translate arbitrary linear threshold circuits into seesaw circuits, we develop four transformation rules: complementation, expansion, consolidation and reduction (see Supplementary Fig. 8 for details). Complementation is used to convert a linear threshold circuit with negative weights into an equivalent circuit with positive weights only

circuit for the general linear threshold gate. Note that thresholds and weights in the final construction are adjusted to obtain improved experimental performance. Fluorophores ROX, FAM, TYE563 and TYE665 are used for four reporters to monitor four outputs  $y_1$  to  $y_4$ . **f**, Kinetics experiments of the general linear threshold gate. A total of 60 DNA strands assembled to form 38 initial DNA species (as indicated by the red numbers in **e**) were mixed in solution at their respective concentrations. The standard concentration was  $1 \times = 16.67$  nM. Input strands  $x_1$  to  $x_3$  were then added with relative concentrations of  $0.1 \times$  (0, logic 'off') or  $0.9 \times$  (1, logic 'on'). Output signals  $y_1$  to  $y_4$  were reported by four distinct fluorophores simultaneously (Supplementary Fig. 2). Trajectories for corresponding inputs are shown with matching colours. Domains and strand sequences are listed in Supplementary Tables 1–4, circuit 2. Experiments were performed at  $20^\circ\text{C}$  in Tris-acetate-EDTA buffer containing  $12.5$  mM  $\text{Mg}^{2+}$ . Output signals were inferred by fluorescence signals normalized to the maximum completion level (Supplementary Fig. 3).

(for example, Fig. 2a). Expansion is used to transform each  $n$ -input linear threshold gate with positive weights into an equivalent network of  $n + 2$  seesaw gates (for example, Fig. 1c). Consolidation is used to collect multiple occurrences of the same signal into one when connecting subcircuits together (for example, yielding nodes 3, 9 and 18 in Fig. 1e). Reduction is used to combine an upstream threshold and a directly downstream weight into a single operation after composition (for example, yielding nodes 17 and 8 in Fig. 2b).

For our initial experimental demonstration, we chose a general three-input four-output linear threshold gate (Fig. 1d). It is equivalent to a linear threshold circuit with four parallel gates that each read the same three inputs. The circuit calculates the analogue value of a three-bit binary number, then compares it to 1, 3, 5 and 7. An equivalent seesaw circuit (Fig. 1e) is generated using just two of the above transformation rules: expansion and consolidation. The first layer of seesaw gates fans out each input while multiplying by the corresponding weight; the second layer calculates the sum of all weighted inputs; the third layer implements the corresponding threshold for each output; the final layer of reporters reads the output signals and provides irreversible drains.

In fluorescence kinetics experiments (Fig. 1f), all four outputs ( $y_1$ – $y_4$ ) achieved the correct 'on' or 'off' states with the complete eight sets of inputs ( $x_3$ ,  $x_2$ ,  $x_1$ , on right side of graphs, colour coded to match traces), even though the inputs were intentionally 'noisy' ( $0.1 \times$  standard concentration was used for 'off' inputs and  $0.9 \times$  for 'on' inputs). In a subcircuit roughly half the size, we tuned weights and thresholds to show that the same set of DNA molecules can implement different linear threshold functions (Supplementary Figs 9 and 10). Although



**Figure 2 | A linear threshold circuit that computes the three-bit XOR function.** This function is given by  $\text{XOR}(x_1, x_2, x_3) = (x_1 + x_2 + x_3) \bmod 2$ . **a**, A three-bit XOR circuit and its equivalent dual-rail circuit. **b**, An equivalent seesaw circuit. **c**, Kinetics experiments. A total of 68 DNA strands assembled to form 42 initial DNA species (as indicated by the red numbers in **b**) were mixed in solution at their respective concentrations. The standard concentration was

$1 \times = 50 \text{ nM}$ . Six dual-rail input strands were then added with relative concentrations of  $0.9 \times x_i^0$  and  $0.1 \times x_i^1$  (for  $x_i = 0$ , logic 'off'), or  $0.1 \times x_i^0$  and  $0.9 \times x_i^1$  (for  $x_i = 1$ , logic 'on'). Dotted and solid lines indicate dual-rail output being logic 'off' and logic 'on', respectively. Domains and strand sequences are listed in Supplementary Tables 1–4, circuit 3. Experiments were performed at  $25^\circ\text{C}$ .

thresholding, catalysis and integration have been demonstrated previously in seesaw digital logic circuits<sup>17</sup>, this is to our knowledge the first demonstration of variable weights, variable thresholds and linear integration composed together, performing the function of an artificial neuron.

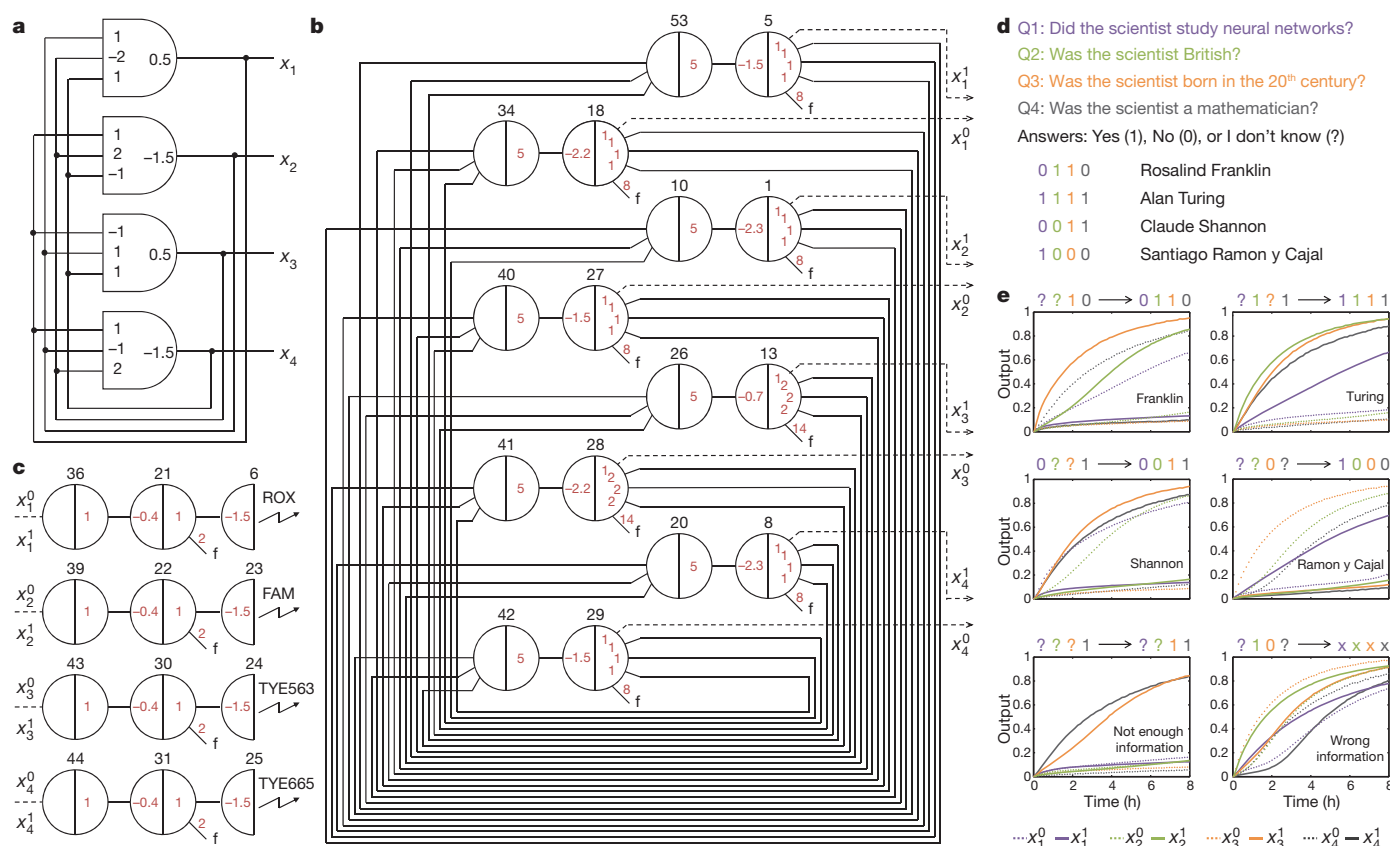
To show computation with negative weights and cascading, a linear threshold circuit that computes the three-bit exclusive-or (XOR) function was demonstrated (Fig. 2a). With an efficient construction, the XOR function with  $n$  variables can be realized with  $\log_2 n$  linear threshold gates<sup>21</sup>, whereas the optimal size of an AND–OR–NOT circuit<sup>22</sup> is at least  $2n$ . All four transformation rules are used to generate an equivalent seesaw circuit (Fig. 2b). Complementation introduces dual-rail logic<sup>24</sup>, where each input  $x_i$  is replaced by a pair of inputs  $x_i^0$  and  $x_i^1$ , representing logic 'off' and logic 'on' separately; each linear threshold gate is replaced by a pair of gates with only positive weights, producing a pair of dual-rail outputs. Thus, a computed 'off' value can be distinguished from an output that has not yet been computed. This method avoids the difficulty of directly implementing negative weights, at a cost of doubling the size of the circuit. The top half of the seesaw circuit corresponds to the cascade of two linear threshold gates that read inputs  $x_i^0$  and output  $y^0$ ; the bottom half corresponds to the other two, which read inputs  $x_i^1$  and output  $y^1$ . The cross-connection between the two halves appears where there is a negative weight in the original circuit.

In fluorescence kinetics experiments (Fig. 2c), the pair of dual-rail outputs went to their correct 'on'/'off' states, again even with noisy inputs. When the inputs had an even number of 1s, the output  $y^0$  went 'on' and  $y^1$  went 'off', indicating  $y = 0$ ; when the inputs had an odd number of 1s, the output  $y^0$  went 'off' and  $y^1$  went 'on', indicating  $y = 1$ . With inputs  $x_1x_2x_3 = 000$  and  $111$ , the output responded sooner than all the other cases, where the production of output must wait for the upstream linear threshold gate to provide its input. Experimental insights gained from the networks of Figs 1 and 2, comparison to a simpler implementation of a three-bit XOR using deoxyribozymes<sup>25</sup>, as well as comparison to other neural network implementations, are discussed in Supplementary Information section 4.

To show a recurrent linear threshold circuit and the power of neural network computation, a four-neuron Hopfield associative memory was demonstrated. A Hopfield network<sup>19</sup> has a number of artificial neurons that are fully connected to each other. If properly trained, which means the weights and threshold of each neuron are properly chosen, the network is able to 'remember' a set of patterns; when initialized with a partial or distorted pattern, the network will recover the most similar remembered pattern. We used the perceptron learning algorithm<sup>1</sup> *in silico* (see Supplementary Information section 5) to train our four-neuron Hopfield network to remember four patterns: 0110, 1111, 0011 and 1000 (Fig. 3a).

To implement negative weights, we again use the dual-rail convention where each signal  $x_i$  is replaced by a pair of signals  $x_i^0$  and  $x_i^1$  (Supplementary Fig. 13). To run the network, there are three possible initial states for each neuron: 0 (logic 'off') if  $x_i^0 = 1$ , 1 (logic 'on') if  $x_i^1 = 1$ , or unknown if both  $x_i^0$  and  $x_i^1 = 0$ . For each update of a linear threshold gate, some  $x_i^0$  or  $x_i^1$  can flip from 0 to 1, but not back; the corresponding neuron can change its state from unknown to 0 or to 1, but not back. As the network runs, another possibility arises: a neuron's state is declared invalid if both  $x_i^0$  and  $x_i^1 = 1$ . Like the original network, this dual-rail Hopfield associative memory can associate an incomplete pattern with a remembered pattern; unlike the original, it is unable to recover a corrupted pattern because the neurons' states cannot flip from 0 to 1 or from 1 to 0; but thanks to the additional states (unknown and invalid), it has the new feature of identifying patterns that are compatible with no single remembered pattern (see Supplementary Information section 5 and the bottom two panels of Fig. 3e).

Following the same transformation rules described above, a seesaw circuit equivalent to the dual-rail Hopfield network is shown in Fig. 3b, containing 24 feedback loops. Initial experiments confirmed that in a circuit with feedback connections between catalytic seesaw gates, leak reactions that occur in DNA strand displacement circuitry are amplified after they exceed the threshold (Supplementary Fig. 14). Therefore, in front of each reporter we add an extra signal restoration step consisting



**Figure 3 | A four-neuron Hopfield associative memory.** **a**, The recurrent linear threshold circuit. **b**, The resulting seesaw circuit using the dual-rail implementation. Dashed lines indicate the connections to reporters. **c**, Four sets of reporters with signal restoration that are connected to either  $x_i^0$  or  $x_i^1$  at any given time. **d**, A 'read your mind' game between a human and the four-neuron DNA associative memory that 'remembers' four scientists according to the answers of four questions. **e**, Kinetics experiments of the 'read your mind' game. A total of 112 DNA strands assembled to form 72 initial DNA species (as indicated by the red numbers in **b**, **c**) were mixed in solution at their respective concentrations. The standard concentration was  $1 \times = 25$  nM. Selected inputs corresponding to the human's answers were then added with relative

concentrations of  $5 \times$  (to set the initial states, inputs triggering the update of multiple neurons are used, for example,  $w_{53,5}$  for  $x_1^1$  and  $w_{34,18}$  for  $x_1^0$ ). Dotted and solid lines indicate dual-rail outputs  $x_i^0$  and  $x_i^1$ , respectively. For each signal, if both dotted and solid lines stay 'off' (less than 0.2), the logic value is unknown, '?'; if the dotted (solid) line goes 'on' (greater than 0.65) and the solid (dotted) line stays 'off', the logic value is '0' ('1'); if both dotted and solid lines go 'on', the logic value is invalid, 'x'. Arrows connect initial states of the four neurons (inputs) to the final states (outputs at 8 h). The eight trajectories in each plot were from two separate experiments (connecting either  $x_i^0$  or  $x_i^1$  to the reporters) because we only have four distinct fluorophores. Sequences of strands are listed in Supplementary Tables 5–7. Experiments were performed at 25 °C.

of an integrating gate and an amplifying gate, in order to suppress leak (Fig. 3c). The values of weights and thresholds determined by *in silico* training were used to determine the concentrations of the 72 DNA species that comprise the memory (Fig. 3b, c). In principle, the same set of DNA molecules could be retrained to remember any of 500 distinct sets of patterns by adjusting weight and threshold concentrations (Supplementary Information section 5).

In the tradition of using game-playing automata as a benchmark for new computing technologies, we demonstrated the Hopfield network in the context of a game called 'read your mind', which is played between a human and the DNA associative memory in a cuvette (Fig. 3d). The game consists of three steps. First, the human thinks of a scientist, choosing from the listed four options (each scientist corresponds to one of the four patterns; for example, Franklin is 0110) or someone else. Second, the human 'tells' the DNA associative memory some of the answers to questions Q1 to Q4 (Fig. 3d) by adding corresponding DNA strands to the cuvette. Finally, after 8 h of 'thinking', the DNA associative memory will guess who is in the human's mind and 'tell' the human the rest of the answers by fluorescence signals. In doing so, the four-neuron DNA associative memory exhibits a brain-like behaviour: associative recall of memories based on incomplete information.

We played the game 27 times with the DNA associative memory, out of 81 possible ways of answering questions Q1 to Q4. Six examples are

shown in Fig. 3e; the rest are shown in Supplementary Figs 15–18. The top left data in Fig. 3e can be interpreted as following: when the human 'said' that the scientist was born in the twentieth century (input  $x_3 = 1$ ) but was not a mathematician (input  $x_4 = 0$ ), the DNA associative memory 'guessed' that the scientist did not study neural networks (output  $x_1$  was updated to 0) but was British (output  $x_2$  was updated to 1), which indicated that the scientist was Rosalind Franklin (pattern 0110). Similarly, the DNA associative memory was able to work out the other three scientists correctly—in the best case, only one answer was given by the human (the middle right data). The bottom left data shows that when the information provided by the human matched multiple patterns (that is, input  $x_4 = 1$  indicates that the scientist was a mathematician, which is true for both Alan Turing and Claude Shannon), the DNA associative memory was able to identify that they were both born in the twentieth century (output  $x_3$  was updated to 1), while the other outputs remained unknown. The bottom right data show that the DNA associative memory was also able to recognize information that was incompatible with all memorized patterns by producing invalid output.

All experiments reported here were semiquantitatively reproduced by mass action simulations using the exact model developed previously for seesaw digital logic circuits<sup>17</sup> with no changes to any rate constants (see Supplementary Information section 7 and Supplementary Figs 19–24 for comparisons to experiments, and Supplementary Figs 25–27 for simulation predictions for the remaining 54 games).



It is interesting to consider the scale of our reactions. Stochastic simulations suggest that the four-neuron DNA associative memory would function reliably with even just 10 copies of each species at  $1 \times$  concentrations (Supplementary Fig. 28), which at our concentrations would entail a volume of roughly  $1 \mu\text{m}^3$ , that is, small enough to fit inside a bacterium. In the other direction, scaling up the DNA associative memory to contain more neurons will exacerbate problems with spurious reactions, and may require lower concentrations and thus slower reactions. On the other hand, neural associative memories are intrinsically fault-tolerant<sup>19</sup> and can function well even with only sparse connections<sup>26</sup>. Because of these opposing factors, it is difficult to predict how large a network can be successfully implemented using the approach described here.

To create smart and functional chemical systems, our current constructions will need to be improved and integrated with other chemistries. For sustained autonomous behaviour, it will be important to go beyond use-once architectures to dynamic units that can turn 'on' and 'off' repeatedly as their inputs change. Initial examples of such systems have been demonstrated using enzymes<sup>13</sup> and are theoretically possible in DNA-only systems<sup>27</sup>. Of particular interest would be to implement the dynamics for learning rules within the chemistry itself<sup>28</sup>, as hinted at by recent demonstrations of trainable chemical circuits<sup>29</sup>. Nonetheless, even simple linear threshold units could be quite useful in biomedical diagnostics, such as classifying cancers with microRNA signals<sup>9,30</sup>. Furthermore, when DNA strand displacement systems are provided with interfaces for sensing non-nucleic-acid inputs and controlling chemical reactions as output actions<sup>31</sup>, an 'intelligent' DNA system could directly perceive and act on its chemical environment.

Received 31 December 2010; accepted 31 May 2011.

- Rojas, R. *Neural Networks: A Systematic Introduction* (Springer, 1996).
- Bray, D. Protein molecules as computational elements in living cells. *Nature* **376**, 307–312 (1995).
- Mjolsness, E., Sharp, D. H. & Reinitz, J. A connectionist model of development. *J. Theor. Biol.* **152**, 429–453 (1991).
- Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl Acad. Sci. USA* **100**, 5136–5141 (2003).
- Hjelmfelt, A., Weinberger, E. D. & Ross, J. Chemical implementation of neural networks and Turing machines. *Proc. Natl Acad. Sci. USA* **88**, 10983–10987 (1991).
- Baum, E. B. Building an associative memory vastly larger than the brain. *Science* **268**, 583–585 (1995).
- Mills, A. P., Yurke, B. & Platzman, P. M. Article for analog vector algebra computation. *Biosystems* **52**, 175–180 (1999).
- Kim, J., Hopfield, J. J. & Winfree, E. in *Advances in Neural Information Processing Systems* Vol. 17 (eds Saul, L. K., Weiss, Y. & Bottou, L.) 681–688 (MIT Press, 2004).
- Zhang, D. Y. & Seelig, G. in *DNA Computing and Molecular Programming* (eds Sakakibara, Y. & Mi, Y.) 176–186 (Lecture Notes in Computer Science, Vol. 6518, Springer, 2011).
- Laplanche, J. P., Pemberton, M., Hjelmfelt, A. & Ross, J. Experiments on pattern recognition by chemical kinetics. *J. Phys. Chem.* **99**, 10063–10065 (1995).
- Mills, A. P. Jr, Turberfield, M., Turberfield, A. J., Yurke, B. & Platzman, P. M. Experimental aspects of DNA neural network computation. *Soft Comput.* **5**, 10–18 (2001).
- Lim, H. W. *et al.* *In vitro* molecular pattern classification via DNA-based weighted-sum operation. *Biosystems* **100**, 1–7 (2010).
- Kim, J. & Winfree, E. Synthetic *in vitro* transcriptional oscillators. *Mol. Syst. Biol.* **7**, 465 (2011).
- Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
- Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chem.* **3**, 103–113 (2011).
- Qian, L. & Winfree, E. A simple DNA gate motif for synthesizing large-scale circuits. *J. R. Soc. Interface* doi:10.1098/rsif.2010.0729 (published online 4 February 2011).
- Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201 (2011).
- Muroga, S. *Threshold Logic and its Applications* Vol. 18 (Wiley-Interscience, 1971).
- Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
- McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biol.* **5**, 115–133 (1943).
- Kautz, W. H. The realization of symmetric switching functions with linear-input logical elements. *IRE Trans. Electron. Comput.* **EC-10**, 371–378 (1961).
- Wegener, I. The complexity of the parity function in unbounded fan-in, unbounded depth circuits. *Theor. Comput. Sci.* **85**, 155–170 (1991).
- Hajnal, A., Maass, W., Pudlák, P., Szegedy, M. & Turan, G. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.* **46**, 129–154 (1993).
- Müller, D. E. in *Symp. on the Application of Switching Theory* (eds Aiken, H. & Main, W. F.) 289–297 (Stanford Univ. Press, 1963).
- Lederman, H., Macdonald, J., Stefanovic, D. & Stojanovic, M. N. Deoxyribozyme-based three-input logic gates and construction of a molecular full adder. *Biochemistry* **45**, 1194–1199 (2006).
- Amari, S. I. Characteristics of sparsely encoded associative memory. *Neural Netw.* **2**, 451–457 (1989).
- Soloveichik, D., Seelig, G. & Winfree, E. DNA as a universal substrate for chemical kinetics. *Proc. Natl Acad. Sci. USA* **107**, 5393–5398 (2010).
- Fernando, C. T. *et al.* Molecular circuits for associative learning in single-celled organisms. *J. R. Soc. Interface* **6**, 463–469 (2009).
- Pei, R., Matamoros, E., Liu, M., Stefanovic, D. & Stojanovic, M. N. Training a molecular automaton to play a game. *Nature Nanotechnol.* **5**, 773–777 (2010).
- Rosenfeld, N. *et al.* MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnol.* **26**, 462–469 (2008).
- Simmel, F. C. Towards biomedical applications for nucleic acid nanodevices. *Nanomedicine* **2**, 817–830 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Rothmund, P. Yin, D. Woods, D. Soloveichik and N. Dabby for comments on the manuscript. We also thank R. Murray for the use of experimental facilities. This work was supported by the NSF (grant nos 0728703 and 0832824 (The Molecular Programming Project)) and by HFSP award no. RGY0074/2006-C.

**Author Contributions** L.Q. designed the system, performed the experiments and analysed the data; L.Q. and E.W. performed the *in silico* training and wrote the manuscript; E.W. guided the project and discussed the design and the data; and J.B. initiated and guided the project, and discussed the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to E.W. ([winfree@caltech.edu](mailto:winfree@caltech.edu)).

# Coseismic and postseismic slip of the 2011 magnitude-9 Tohoku-Oki earthquake

Shinzauro Ozawa<sup>1</sup>, Takuya Nishimura<sup>1</sup>, Hisashi Suito<sup>1</sup>, Tomokazu Kobayashi<sup>1</sup>, Mikio Tobita<sup>1</sup> & Tetsuro Imakiire<sup>1</sup>

Most large earthquakes occur along an oceanic trench, where an oceanic plate subducts beneath a continental plate. Massive earthquakes with a moment magnitude,  $M_w$ , of nine have been known to occur in only a few areas, including Chile, Alaska, Kamchatka and Sumatra. No historical records exist of a  $M_w = 9$  earthquake along the Japan trench, where the Pacific plate subducts beneath the Okhotsk plate, with the possible exception of the AD 869 Jogan earthquake<sup>1</sup>, the magnitude of which has not been well constrained. However, the strain accumulation rate estimated there from recent geodetic observations is much higher than the average strain rate released in previous interplate earthquakes<sup>2–6</sup>. This finding raises the question of how such areas release the accumulated strain. A megathrust earthquake with  $M_w = 9.0$  (hereafter referred to as the Tohoku-Oki earthquake) occurred on 11 March 2011, rupturing the plate boundary off the Pacific coast of northeastern Japan. Here we report the distributions of the coseismic slip and postseismic slip as determined from ground displacement detected using a network based on the Global Positioning System. The coseismic slip area extends approximately 400 km along the Japan trench, matching the area of the pre-seismic locked zone<sup>4</sup>. The afterslip has begun to overlap the coseismic slip area and extends into the surrounding region. In particular, the afterslip area reached a depth of approximately 100 km, with  $M_w = 8.3$ , on 25 March 2011. Because the Tohoku-Oki earthquake released the strain accumulated for several hundred years, the paradox of the strain budget imbalance may be partly resolved. This earthquake reminds us of the potential for  $M_w \approx 9$  earthquakes to occur along other trench systems, even if no past evidence of such events exists. Therefore, it is imperative that strain accumulation be monitored using a space geodetic technique to assess earthquake potential.

Northeastern Japan has been struck by many M7-class ( $M_w = 7$ ) interplate earthquakes along the Japan trench, where the Pacific plate subducts beneath the Okhotsk plate at a rate of 73–78 mm yr<sup>-1</sup> (refs 7, 8; Fig. 1a). However, no interplate earthquake along the Japan trench with a surface-wave magnitude,  $M$ , of more than 7.5 has been instrumentally recorded since 1923, except along the northernmost part of the trench, where there have been  $M = 7.9$  and  $M = 7.6$  earthquakes (Fig. 1b). There are no historical records of any  $M_w > 8.5$  earthquakes occurring along the Japan trench since the seventeenth century. Therefore, the massive ( $M_w = 9$ ) Tohoku-Oki earthquake was not widely anticipated despite the geological evidence for recurrent devastating tsunamis in the past, in particular in AD 869<sup>1</sup>, and the rapid accumulation of elastic strain along the trench.

On the basis of earthquake catalogues covering several decades<sup>9</sup>, the seismic coupling coefficient, that is, the ratio between the rate of slip released in an interplate earthquake and the rate of relative plate motion, has been estimated to be 10–20% along the Japan trench<sup>2,3</sup>. However, ground displacement data acquired using a continuous Global Positioning System (GPS) network established in 1994 suggest that there is strong interplate coupling along the Japan trench<sup>4–6</sup> (Fig. 1b). The strain accumulation rate estimated from contemporary deformation is considerably higher than the average strain rate

released in historical earthquakes. An episodic aseismic slip, including an afterslip, has been suggested as a possible mechanism for significant elastic strain release<sup>4,10</sup>.

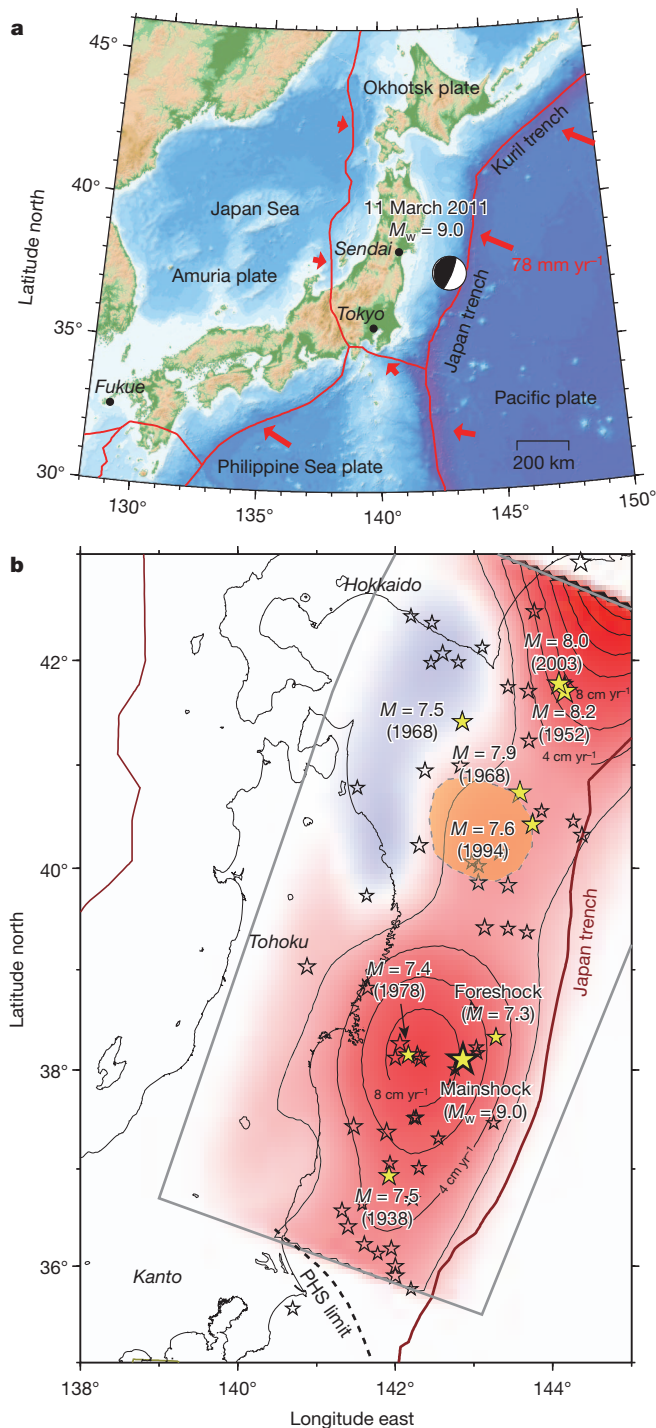
In this Letter, first we describe the coseismic and postseismic deformations associated with the Tohoku-Oki earthquake, as detected by the GPS Earth Observation Network (GEONET) operated by the Geospatial Information Authority of Japan<sup>11</sup>, and estimate the coseismic slip distribution and the subsequent afterslip distribution on the plate boundary by geodetic inversion<sup>12</sup> of the ground displacement at selected GPS sites. Second, we discuss the relationship between the coseismic and the postseismic slip models, as well as their relation to pre-seismic coupling and strain budget imbalance.

The observed coseismic displacements show eastward movements of up to 5.3 m and subsidence by up to 1.2 m along the coastal line of the Tohoku region, relative to the Fukue site (Figs 1 and 2a). These values are greater by one order of magnitude than those recorded at the time of previous M7–M8-class interplate earthquakes in the Tohoku region that have taken place since the establishment of GEONET. After the Tohoku-Oki earthquake, a large postseismic deformation occurred (Fig. 2b). Although the postseismic deformation resembles the coseismic field, the displacements seem to be more broadly distributed. In particular, the eastward displacement of the Pacific coastal area did not differ significantly from that of the western coastal area, whereas the eastward displacement on the Pacific coast was much larger than that of the western coastal area in the coseismic field. In addition, the Pacific coastal area near the source region was uplifted after the earthquake.

The slip distribution estimated on the basis of the coseismic displacements shows a large slip of up to 27 m near the epicentral area, extending approximately 400 km along the Japan trench at a depth of less than 60 km, which is the lower limit for the seismogenic zone along the subducting plate in this region<sup>13</sup> (Figs 1b and 2a). The estimated moment is  $3.43 \times 10^{22}$  N m, assuming a uniform rigidity of 40 GPa, equivalent to that of an  $M_w = 9.0$  earthquake. A uniform rigidity of 40 GPa is a rough average of 29, 41 and 50 GPa for the typical rigidities of upper crust, lower crust and upper mantle in northeastern Japan based on seismic data<sup>14</sup>. The moment of our geodetic model closely matches the moment magnitude of 9.1 inferred from seismic waveform analysis in ref. 15. The root mean squared deviation of this model is 0.011 m (Supplementary Figs 1 and 2), and the estimated slip is well beyond the  $1\sigma$  error (Supplementary Figs 3). Chequerboard and sensitivity tests show that spatial variations of slip over the coseismic area are resolved to the scale of a few tens of kilometres and that the principal pattern is a stable feature.

The estimated afterslip, which is based on the postseismic deformation, occurs in the coseismic slip area and adjacent to it, expanding to the north, the south and in the dipping direction (Figs 2b and 3). The afterslip area has two modal centres: northwest of the centre of the coseismic slip and east of the Kanto region. These centres reflect the large postseismic displacement along the Pacific coast, which, unlike the coseismic displacement, extends north and south (Fig. 2). The root mean squared deviation of this model is 0.007 m (Supplementary Figs 3 and 4). The estimated moment of the afterslip on 25 March is

<sup>1</sup>Geospatial Information Authority of Japan, Tsukuba, Ibaraki 305-0811, Japan.



**Figure 1 | Tectonic setting in and around the Tohoku-Oki earthquake.**

**a**, Plate configurations of the Japanese islands<sup>29</sup>. The focal mechanism of the Tohoku-Oki earthquake is taken from the Global Centroid-Moment-Tensor Project<sup>15</sup>. The red arrows indicate relative motion between the two plates at a plate boundary<sup>7,8</sup>. **b**, Coupling distribution before the earthquake and recent seismicity along the Japan trench. The colour shading and contours indicate the degree of interplate coupling between the subducting Pacific plate and the overriding Okhotsk plate, estimated from GPS data recorded between April 2000 and March 2001<sup>4</sup>. The degree of coupling is expressed as the backslip rate<sup>30</sup>, which is a slip deficit from the relative plate velocity. The stars mark the epicentres of large ( $M \geq 6.8$ ) earthquakes that have occurred since 1923. The epicentres of the mainshock, a foreshock and earthquakes with  $M \geq 7.4$  are marked by yellow stars and labelled with their magnitudes and/or times of occurrence. The orange area is the source area of the  $M = 7.6$  1994 earthquake<sup>20</sup>. The dashed line shows the northeastern limit of the subducted Philippine Sea plate<sup>21</sup> (PHS). The Okhotsk plate overrides the Pacific plate north of this limit and the Philippine Sea plate overrides the Pacific plate south of this limit. The grey rectangle represents a fault patch to estimate the backslip rate.

afterslip area, avoiding the large coseismic slip area (Fig. 3). This is consistent with the observation that in many cases the aftershocks and the large afterslip occur in an area where the coseismic slip is not large<sup>18,19</sup>. The seismic moments of thrust-type aftershocks sum to  $1.5 \times 10^{19}$  N m for the period of the postseismic deformation. This suggests that aftershocks contribute less than 1% of the moment of the afterslip model for two weeks. Although the estimated area of small afterslip overlaps the coseismic slip, we cannot rule out the possibility that this may be due to oversmoothing in the afterslip estimation, because a sensitivity test of the smoothing constraint shows that the afterslip area avoids the large coseismic slip area in undersmoothed models. The expansion in the dipping direction reaches 80–100 km in depth, which is the lower limit for the coupling of the plates in this region.

The propagation of the afterslip into an area deeper than the coseismic slip area was observed for the 1994 Sanriku-Haruka-Oki earthquake<sup>20</sup> ( $M = 7.6$ ) and the 2003 Tokachi-Oki earthquake<sup>18</sup> ( $M = 8.0$ ), suggesting a general dipping expansion along the Japan and Kuril trenches. Because the interplate coupling rate is near zero at depths of more than 100 km, we think that the afterslip area terminates at this limit.

The northward expansion of the afterslip area approaches the zone ruptured in the 1968 ( $M = 7.9$ ) and 1994 ( $M = 7.6$ ) earthquakes, (Fig. 3). The afterslip area may terminate there because the source area of the 1994 earthquake is now strongly locked. The southward expansion has reached the Kanto region (Figs 1b, 2b and 3). The Philippine Sea plate overrides the Pacific plate in the Kanto region, south of the northeastern limit of the former plate, whereas the Okhotsk plate lies on the Pacific plate north of this limit<sup>21</sup> (Figs 1 and 3). There is a possibility that this change in the overriding plate stops the southward expansion of the afterslip at the limit of the Philippine Sea plate in the Kanto region. Our model estimates the afterslip distribution at intervals of 1 d, and the two modal centres of the afterslip area seem not to move significantly on this timescale whereas the slip magnitude increases rapidly.

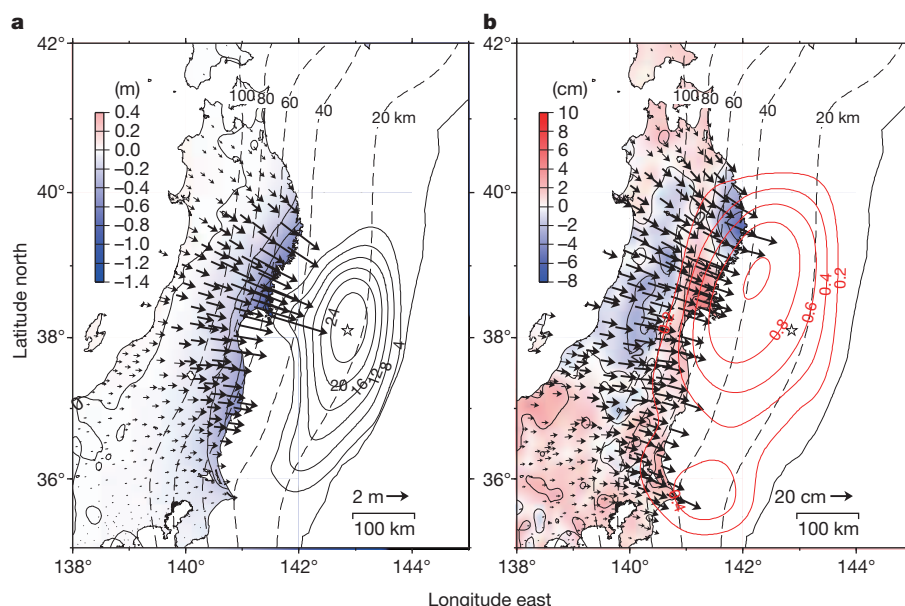
The ruptured area of the Tohoku-Oki earthquake well matches the area estimated to have been strongly coupled before the earthquake<sup>4</sup> (Fig. 1b), although the centre of the coseismic slip area is shallower than that of the locked area. This finding indicates the extreme importance of the GPS observations in assessing the potential of a subduction earthquake occurring, as has been observed in other subduction zones<sup>22</sup>. A deeper part of the locked zone may release the remaining strain energy by further afterslip of the Tohoku-Oki earthquake.

The moment accumulation rate attributed to the subduction of the Pacific plate in an area from latitude  $36^\circ$  N to  $39.5^\circ$  N along the Japan trench is estimated to have been approximately  $1.6 \times 10^{20}$  N m yr<sup>-1</sup> before the earthquake<sup>4</sup>. The repeated occurrence of  $M_w < 8$  interplate earthquakes contributes 10–20% of the plate motion, without taking the afterslip into account<sup>2,3</sup>. It remains unclear how much energy will be released by the afterslip of the Tohoku-Oki earthquake. In the case

$3.35 \times 10^{21}$  N m, which is equivalent to that of an  $M_w = 8.3$  earthquake (Figs 2b and 3). This moment is approximately 10% of that of the mainshock. We assume that the postseismic deformation transients are due solely to afterslip, although they are affected by the viscoelastic relaxation of the asthenosphere and poroelastic rebound<sup>16</sup>. We estimate the magnitude of these effects by simple calculation. The viscoelastic relaxation model<sup>17</sup> is predicted to be within 1 cm of surface displacement for two weeks, with an asthenospheric viscosity of  $10^{19}$  Pa. Although poroelastic effects reach 20% of the observed post-seismic deformation at maximum, their horizontal and vertical patterns are quite different from the observations. Thus, we assume that these effects can be ignored as a first approximation in this case.

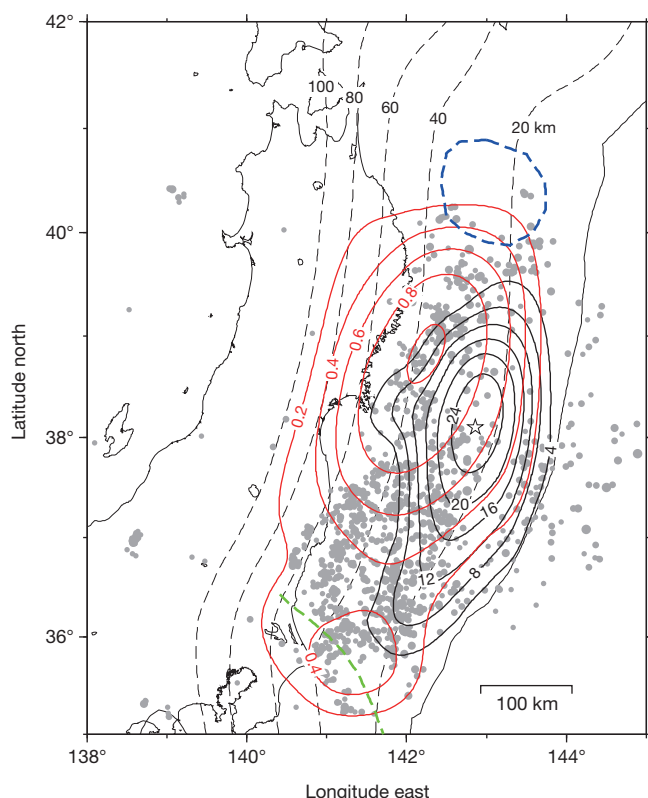
The area of large afterslip is located in the region peripheral to the coseismic slip zone. In addition, aftershocks seem to occur in the





**Figure 2 | Coseismic and postseismic displacements and estimated slip.** **a**, Coseismic displacements for 10–11 March 2011, relative to the Fukue site. The black arrows indicate the horizontal coseismic movements of the GPS sites. The colour shading indicates vertical displacement. The star marks the location of the earthquake epicentre. The dotted lines indicate the isodepth contours of

the plate boundary at 20-km intervals<sup>28</sup>. The solid contours show the coseismic slip distribution in metres. **b**, Postseismic displacements for 12–25 March 2011, relative to the Fukue site. The red contours show the afterslip distribution in metres. All other markings represent the same as in **a**.



**Figure 3 | Coseismic slip, postseismic slip and aftershocks.** Estimated coseismic slip (black contour, 4-m interval) and postseismic slip (red contour, 0.2-m interval) of the Tohoku-Oki earthquake for the same period as in Fig. 2. The green dashed line indicates the northeast limit of the Philippine Sea plate. The blue dashed line indicates the ruptured area of the  $M = 7.6$  1994 earthquake<sup>20</sup>. The grey circles show the epicentres of the aftershocks of the Tohoku-Oki earthquake for 11–25 March 2011. All other markings represent the same as in Fig. 2.

of the 1994 Sanriku-Haruka-Oki earthquake, the afterslip released moment equivalent to that of the mainshock over the course of 1 yr (ref. 10). In other cases, it has been reported that the postseismic slip releases a moment of up to  $\sim 30\%$  of that of the mainshock over several weeks after M8-class earthquakes<sup>23</sup>. The moment released by afterslip after the 2004 Sumatra earthquake is also estimated to have been 30% of the moment of the mainshock over 40 d (ref. 24). If we assume that the afterslip of the Tohoku-Oki earthquake eventually releases 30–100% of the amount of energy released by the mainshock, and extrapolate the interplate coupling from the geodetic observations, we find that it would take approximately 350–700 yr for energy equivalent to that of the earthquake to accumulate along the Japan trench.

Recent geological studies suggest that tsunamis similar to that which followed the Tohoku-Oki earthquake have repeatedly struck the Pacific coast of northeastern Japan, with a recurrence interval of approximately 800–1,100 yr (ref. 1), implying that megathrust earthquakes have also occurred repeatedly along the Japan trench. The massive Tohoku-Oki earthquake supports this hypothesis and may partly resolve the strain budget imbalance, although the roughly estimated recurrence interval is shorter than the tsunami return period.

The Pacific coastal area subsided by up to 1.2 m following the earthquake. Furthermore, subsidence of the Pacific coast at a rate of  $5\text{--}10\text{ mm yr}^{-1}$  over the past 100 yr has been estimated by tide gauges<sup>25</sup>, levelling and GPS data. Although geodetic observations indicate both coseismic and interseismic subsidence, a geomorphological study has shown that there was long-term upheaval along the Pacific coast in the late Quaternary period<sup>26</sup>. This discrepancy suggests the existence of another mechanism of episodic uplift, such as postseismic deformation<sup>27</sup>. In fact, the Pacific coastal area near the epicentre started to uplift after the Tohoku-Oki earthquake by an amount ranging from 1 to 4 cm, as observed over the course of two weeks (Fig. 2). It is difficult to predict the future temporal evolution of uplifting from such a short observation period. If the uplift lasts for a long time, the discrepancy between subsidence and upheaval will be resolved. To understand the uplift mechanism, it is important to continue geodetic monitoring.

## METHODS SUMMARY

Coseismic and postseismic displacements were based on GPS data collected for 6 h and were analysed with the BERNESE GPS software. We used the east–west,

north–south and up–down components, measured relative to the Fukue site, at approximately 400 selected GPS sites covering northeastern Japan (Fig. 1a). We used the Yabuki–Matsu'ura method<sup>12</sup> to estimate the slip distribution on the plate boundary, and used a fault patch that covers an area ~500 km in width and ~800 km in length to represent the plate boundary in this region<sup>28</sup>. The fault patch was represented by a parametric spline surface. Green's functions<sup>12</sup> that assume a homogeneous half-space were used. A detailed description of the GPS inversion approach, including resolution and sensitivity analysis, can be found in Methods and Supplementary Figs 5–9. The data set and the results of inversion are shown in Supplementary Tables 1–4.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 31 March; accepted 25 May 2011.**

**Published online 15 June 2011.**

- Minoura, K., Imamura, F., Sugawara, D., Kono, Y. & Iwashita, T. The 869 Jogan tsunami deposit and recurrence interval of large-scale tsunami on the Pacific coast of northeast Japan. *J. Nat. Disaster Sci.* **23**, 83–88 (2001).
- Peterson, E. T. & Seno, T. Factors affecting seismic moment release rates in subduction zones. *J. Geophys. Res.* **89**, 10233–10248 (1984).
- Pacheco, J. F., Sykes, L. R. & Scholz, C. H. Nature of seismic coupling along simple plate boundaries of the subduction type. *J. Geophys. Res.* **98**, 14133–14159 (1993).
- Nishimura, T. *et al.* Temporal change of interplate coupling in northeastern Japan during 1995–2002 estimated from continuous GPS observations. *Geophys. J. Int.* **157**, 901–916 (2004).
- Hashimoto, C., Noda, A., Sagiya, T. & Matsu'ura, M. Interplate seismogenic zones along the Kuril–Japan trench from GPS data inversion. *Nature Geosci.* **2**, 141–144 (2009).
- Suwa, Y., Miura, S., Hasegawa, A., Sato, T. & Tachibana, K. Interplate coupling beneath NE Japan inferred from three-dimensional displacement field. *J. Geophys. Res.* **111**, B04402 (2006).
- Sella, G. F., Dixon, T. H. & Mao, A. REVEL: a model for recent plate velocities from space geodesy. *J. Geophys. Res.* **107**, 2081 (2002).
- Apel, E. V. *et al.* Independent active microplate tectonics of northeast Asia from GPS velocities and block modeling. *Geophys. Res. Lett.* **33**, L11303 (2006).
- Utsu, T. Chronological table of earthquakes in Japan with a moment magnitude larger than 6.0 and disastrous earthquakes from 1885 to 1980. *Bull. Earthq. Res. Inst.* **57**, 401–463 (1983).
- Heki, K., Miyazaki, S. & Tsuji, H. Silent fault slip following an interplate thrust earthquake at the Japan Trench. *Nature* **386**, 595–597 (1997).
- Sagiya, T., Miyazaki, S. & Tada, T. Continuous GPS array and present-day crustal deformation of Japan. *Pure Appl. Geophys.* **157**, 2303–2322 (2000).
- Yabuki, T. & Matsu'ura, M. Geodetic data inversion using a Bayesian information criterion for spatial distribution of fault slip. *Geophys. J. Int.* **109**, 363–375 (1992).
- Igarashi, T., Matsuzawa, T., Umino, N. & Hasegawa, A. Spatial distribution of focal mechanisms for and intraplate earthquakes associated with the subducting Pacific plate beneath the northeastern Japan arc: a triple-planed deep seismic zone. *J. Geophys. Res.* **106**, 2177–2191 (2001).
- Nakajima, J. & Matsuzawa, T. Hasegawa, A. & Zhao, D. Seismic imaging of arc magma and fluids under the central part of northeast Japan. *Tectonophysics* **341**, 1–17 (2001).
- Global CMT Web Page. *Global Centroid Moment Tensor Project* (<http://www.globalcmt.org/>) (accessed 30 March 2011).
- Shearer, P. & Burgmann, R. Lessons learned from the 2004 Sumatra–Andaman megathrust rupture. *Annu. Rev. Earth Planet. Sci.* **38**, 103–131 (2010).
- Pollitz, F. Gravitational viscoelastic postseismic relaxation on a layered spherical earth. *J. Geophys. Res.* **102**, 17921–17941 (1997).
- Ozawa, S., Kaidzu, M., Murakami, M., Imakiire, T. & Hatanaka, Y. Coseismic and postseismic crustal deformation after the  $M_w$  8 Tokachi-oki earthquake in Japan. *Earth Planets Space* **56**, 675–680 (2004).
- Hsu, Y. *et al.* Frictional afterslip following the 2005 Nias–Simeulue earthquake, Sumatra. *Science* **312**, 1921–1926 (2006).
- Nishimura, T. *et al.* Distribution of seismic coupling on the subducting plate boundary in northeastern Japan inferred from GPS observations. *Tectonophysics* **323**, 217–238 (2000).
- Uchida, N., Matsuzawa, T., Nakajima, J. & Hasegawa, A. Subduction of a wedge-shaped Philippine Sea plate beneath Kanto, central Japan, estimated from converted waves and small repeating earthquakes. *J. Geophys. Res.* **115**, B07309 (2010).
- Moreno, M., Rosenau, M. & Oncken, O. 2010 Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone. *Nature* **467**, 198–202 (2010).
- Melbourn, T. I., Webb, F. H., Stock, J. M. & Reigbar, C. Rapid postseismic transients in subduction zones from continuous GPS. *J. Geophys. Res.* **107**, 2241 (2002).
- Chlieh, M. *et al.* Coseismic slip and afterslip of the great  $M_w$  9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97**, S152–S173 (2007).
- Kato, T. Secular and earthquake-related vertical crustal movements in Japan as deduced from tidal records (1951–1981). *Tectonophysics* **97**, 183–200 (1983).
- Matsu'ura, T., Furusawa, A. & Saomoto, H. Long-term and short-term vertical velocity profiles across the forearc in the NE Japan subduction zone. *Quat. Res.* **71**, 227–238 (2009).
- Suito, H. & Freymueller, J. T. A viscoelastic and afterslip postseismic deformation model for the 1964 Alaska earthquake. *J. Geophys. Res.* **114**, B11404 (2009).
- Nakajima, J. & Hasegawa, A. Anomalous low-velocity zone and linear alignment of seismicity along it in the subducted Pacific slab beneath Kanto, Japan: reactivation of subducted fracture zone? *Geophys. Res. Lett.* **33**, L16309 (2006).
- Bird, P. An updated digital model of plate boundaries. *Geochem. Geophys. Geosyst.* **4**, 1027 (2003).
- Savage, J. C. A dislocation model of strain accumulation and release at a subduction zone. *J. Geophys. Res.* **88**, 4984–4996 (1983).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to the Japan Meteorological Agency for providing us with the hypocentre data. We also thank M. Murakami, H. Munekane and Y. Hatanaka for their comments.

**Author Contributions** S.O., T.N. and H.S. participated in the construction of the model and wrote the manuscript. M.T., T.J. and T.K. participated in the discussion of the results and reviewed the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.O. ([ozawa@gsi.go.jp](mailto:ozawa@gsi.go.jp)).

## METHODS

The objective of our study was to determine the distribution of the coseismic and postseismic slip of the Tohoku-Oki earthquake, Japan. The analysis was based on the modelling of the coseismic and postseismic deformation by a GPS network in Japan. A solution was obtained by inverting a set of 377 coseismic slip GPS vectors and 357 postseismic slip GPS vectors along the Japan trench.

**GPS data.** The GPS data used in this study were derived by the most rapid of the three strategies in the baseline analysis of the GEONET routine<sup>31,32</sup>. In the strategy for this solution, 6-h data are processed to estimate the static coordinates using the software BERNES 5.0 and the ultrarapid ephemerides of the International Global Navigation Satellite Systems Service (IGS) every 3 h. Tropospheric zenith delays and gradients for two and, respectively, a single time segment were estimated using the Niell mapping function<sup>33</sup> in each session. The elevation cut-off angle for the GPS satellites was 15°, and absolute phase centre corrections to the GPS antennas were applied for each monument type of the GEONET stations. Thus, the GPS carrier-phase ambiguities were resolved. The site coordinates were compared with the 2005 International Terrestrial Reference Frame<sup>34</sup> (ITRF2005) by fixing a fiducial station (station 92110, close to the Tsukuba IGS station) as the a-priori value using ITRF2005. Because of the large coseismic displacement of the Tohoku-Oki earthquake, the fiducial station moved eastwards by ~0.5 m. In the baseline analysis after the earthquake, the a-priori coordinate of the fiducial station was corrected by adding the coseismic offset. The coseismic and postseismic displacements were calculated from the relative site coordinates with respect to the reference site (Fukue; station code 950462), which is ~1,400 km away from the earthquake epicentres. The repeatabilities of the relative coordinates were 4.0, 3.6 and 15.1 mm for the east–west, north–south and up–down components, respectively, which are averages of standard deviations for the 412 sites used during 1–10 February 2011.

**Model strategy.** Coseismic offsets were estimated by subtracting the average coordinates for the period between 21:00 on 9 March and 09:00 on 11 March from the coordinates at 18:00 on 11 March 2011 (Japan Standard Time). According to the repeatabilities of the coordinates, the errors in the coseismic offsets were approximately 6 and 20 mm in the horizontal and vertical components, respectively. Postseismic data on 25 March were estimated by subtracting coordinates at 18:00 on 11 March from those at 18:00 on 25 March. We used the same error estimates as those for the coseismic deformation. We used 377 and 357 GPS sites in inversion for coseismic slip and postseismic slip, respectively.

We created a parametric spline surface to represent the surface of the subducting Pacific plate in the Tohoku region<sup>28,35</sup>. The parametric spline surface consisted of 15 knots in the dipping direction and 25 knots along the Japan trench. An adopted spline surface covers approximately 500 km in width, 800 km in length and 200 km in depth.

We used an inversion method<sup>12</sup> with minor modifications. In our inversion method, east–west and north–south slip components are represented by a parametric spline surface, as is the case for a fault patch. The vertical slip component is estimated using the formula in ref. 12. Green's functions are calculated using the formulation of ref. 12, which assumes a homogeneous isotropic half-space. We included a roughness matrix<sup>12</sup> as prior information, imposing the condition  $\lambda_1 \mathbf{M} \mathbf{u} = \mathbf{0}$  on the inversion equation, where  $\lambda_1$  is a hyperparameter of roughness,  $\mathbf{M}$  is the roughness matrix<sup>12</sup> and  $\mathbf{u}$  represents slip. We also adopted the prior information that the slip direction is parallel to the motion of the subducting Pacific plate, by imposing the condition  $\lambda_2 (\mathbf{u}_1 - \mathbf{u}_2) = \mathbf{0}$  on the inversion equation, where  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are slip vectors angled at 45° relative to the direction of plate motion and  $\lambda_2$  is a hyperparameter. These roughness and slip rake constraints result in a stable solution. By including the roughness and slip directions, we were able to estimate the two hyperparameters by minimizing the Akaike Bayesian information criterion<sup>12,36</sup>. Minimization was done using Powell's method<sup>37</sup>. We set to zero the displacements at the edge of the fault surface and the displacements east of the Japan trench, as a boundary condition.

**Resolution and sensitivity of the inversion.** We tested the resolution power by attempting to recover a given coseismic slip distribution, which is often called

a chequerboard resolution test. We discretized the fault plane into regular chequerboard patterns with patches assigned either 0 or 5 m eastward interplate slip for the coseismic case (Supplementary Fig. 5a). A forward model introduces the chequerboard pattern as a fault slip condition and then simulates the displacements at the GPS sites. Gaussian random noises corresponding to GPS measurement errors were then added to obtain a set of synthetic data, which we subsequently inverted. The initial chequerboard pattern used in the test had an approximate size of 100 km × 100 km, which dimensions are similar to the wavelength of the structures we want to resolve, that is, the coseismic slip. The preferred models in the coseismic case recovered the main features of the chequerboard in most parts of the model region (Supplementary Fig. 5b). Resolution was generally good in the down-dip and onshore regions but was relatively poor near the trench. A chequerboard pattern was almost reproduced in the western part of the large coseismic slip area. This indicates that our inversion method was able to recover the main features of the coseismic slip distribution in the Tohoku-Oki earthquake.

By using a set of roughness coefficients, ranging from the oversmoothed (Supplementary Fig. 6a–c) to the undersmoothed (Supplementary Fig. 6e–f), we tested the sensitivity of the slip distribution to the roughness coefficient,  $\lambda_1$ . Because  $\lambda_2$ , which constrains the slip direction, does not significantly affect the slip distribution, we show the sensitivity only for  $\lambda_1$ . Supplementary Fig. 6d shows the optimal slip model, as determined using the minimum Akaike Bayesian information criterion. This sensitivity test suggests that the estimated coseismic slip distribution characterized by the area of large slip east of Sendai and the ~200–400-km-long slip area along the Japan trench is a robust feature that does not depend on the chosen roughness of the slip distribution.

We conducted similar tests for the postseismic slip. Supplementary Fig. 7a shows the initial chequerboard pattern, in which we assigned 0.4 and 0 m eastward interplate slip to the pink and blue areas, respectively. The resulting model shows good recovery of the chequerboard pattern, especially in the large afterslip area of Figs 2 and 3, but it suggests relatively poor recovery in the offshore area, which is 100 km away from land. Supplementary Fig. 8d depicts the optimal slip model. This sensitivity test indicates that the centre of the afterslip area is located in the down-dip area of the coseismic slip region, which is independent of the roughness constraint. However, the undersmoothed model (Supplementary Fig. 8e, f) shows that the afterslip concentrates along the down-dip edge of the large coseismic area.

We also tested a situation in which the afterslip does not overlap the coseismic slip area, and checked whether the afterslip distribution assigned in this case was reproduced by our inversion. The assumed condition is the same as the chequerboard test for the postseismic case, except for a given slip distribution. The results show that the assigned slip is well recovered by inversion, although the area of small slip is extended into the coseismic area owing to the smoothness constraint (Supplementary Fig. 9). Thus, we conclude that the slip centre shifts to a deeper portion of the coseismic rupture area in the postseismic period. However, owing to the limited resolving power of the geodetic data, it is not clear whether the afterslip area overlaps the coseismic slip area.

31. Hatanaka, Y. *et al.* Improvement of the analysis strategy of GEONET. *Bull. Geogr. Surv. Inst.* **49**, 11–37 (2003).
32. Nakagawa, H. Development and validation of GEONET new analysis strategy (version 4) [in Japanese]. *J. Geogr. Surv. Inst.* **118**, 1–8 (2009).
33. Niell, A. E. Global mapping functions for the atmosphere delay at radio wavelengths. *J. Geophys. Res.* **101**, 3227–3246 (1996).
34. Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B. & Boucher, C. ITRF2005: a new release of the International Terrestrial Reference Frame based on time series of station positions and Earth orientation parameters. *J. Geophys. Res.* **112**, B09401 (2007).
35. Ozawa, S., Murakami, M. & Tada, T. Time-dependent inversion study of the slow thrust event in the Nankai trough subduction zone, southwestern Japan. *J. Geophys. Res.* **106**, 787–802 (2001).
36. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **AC-19**, 716–723 (1974).
37. Press, H. W., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in Fortran* 409–411 (Cambridge Univ. Press, 1992).



# Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking

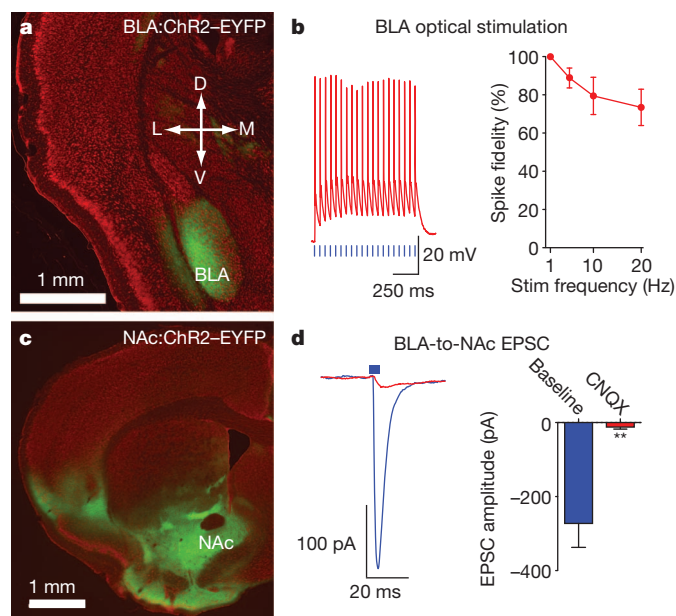
Garret D. Stuber<sup>1,2</sup>, Dennis R. Sparta<sup>1,2</sup>, Alice M. Stamatakis<sup>1</sup>, Wieke A. van Leeuwen<sup>2</sup>, Juanita E. Hardjoprajitno<sup>2</sup>, Saemi Cho<sup>2</sup>, Kay M. Tye<sup>2,3</sup>, Kimberly A. Kempadoo<sup>2</sup>, Feng Zhang<sup>3</sup>, Karl Deisseroth<sup>3</sup> & Antonello Bonci<sup>2,4</sup>

The basolateral amygdala (BLA) has a crucial role in emotional learning irrespective of valence<sup>1–5,21–23</sup>. The BLA projection to the nucleus accumbens (NAc) is thought to modulate cue-triggered motivated behaviours<sup>4,6,7,24,25</sup>, but our understanding of the interaction between these two brain regions has been limited by the inability to manipulate neural-circuit elements of this pathway selectively during behaviour. To circumvent this limitation, we used *in vivo* optogenetic stimulation or inhibition of glutamatergic fibres from the BLA to the NAc, coupled with intracranial pharmacology and *ex vivo* electrophysiology. Here we show that optical stimulation of the pathway from the BLA to the NAc in mice reinforces behavioural responding to earn additional optical stimulation of these synaptic inputs. Optical stimulation of these glutamatergic fibres required intra-NAc dopamine D1-type receptor signalling, but not D2-type receptor signalling. Brief optical inhibition of fibres from the BLA to the NAc reduced cue-evoked intake of sucrose, demonstrating an important role of this specific pathway in controlling naturally occurring reward-related behaviour. Moreover, although optical stimulation of glutamatergic fibres from the medial prefrontal cortex to the NAc also elicited reliable excitatory synaptic responses, optical self-stimulation behaviour was not observed by activation of this pathway. These data indicate that whereas the BLA is important for processing both positive and negative affect, the glutamatergic pathway from the BLA to the NAc, in conjunction with dopamine signalling in the NAc, promotes motivated behavioural responding. Thus, optogenetic manipulation of anatomically distinct synaptic inputs to the NAc reveals functionally distinct properties of these inputs in controlling reward-seeking behaviours.

To stimulate excitatory fibres projecting from the BLA to the NAc selectively, we stereotactically delivered adeno-associated viral vectors carrying the codon-optimized *channelrhodopsin-2* gene fused in-frame to enhanced yellow fluorescent protein (*ChR2-EYFP*)<sup>8</sup>, driven by the *Camk2α* promoter, to transduce glutamatergic neurons locally in the BLA. Expression of ChR2-EYFP was observed after transduction of neurons in the BLA (Fig. 1a). Whole-cell recordings from visually identified BLA pyramidal neurons expressing ChR2 showed that light stimulation frequencies (1–20 Hz, 5-ms light pulses) resulted in reliable firing in response to light, with minimal loss of spike fidelity at 20 Hz (Fig. 1b and Supplementary Fig. 1). This indicated that optically induced firing via activation of ChR2 can excite BLA neurons at physiologically relevant frequencies<sup>5,6</sup>. Expression of ChR2-EYFP was observed in targets of the BLA in the forebrain, including the NAc (Fig. 1c). Optical stimulation of ChR2-EYFP-positive fibres and synaptic terminals from the BLA to the NAc resulted in excitatory responses in the NAc (Fig. 1d and Supplementary Fig. 2). Light-evoked excitatory postsynaptic currents (EPSCs) from visually identified medium spiny neurons were blocked by bath application of the

competitive  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA) antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) at 10  $\mu$ M, demonstrating that optical stimulation of BLA-to-NAc fibres results in AMPAR-mediated EPSCs via the release of synaptic glutamate (Fig. 1d).

To test whether selective activation of BLA-to-NAc synapses could promote motivated behavioural responding, mice injected into the BLA with viruses encoding ChR2-EYFP or EYFP alone (control) were stereotactically implanted with a guide cannula above the ipsilateral NAc. At 21–28 d after surgery, a fibre-optic cable connected to a laser capable of activating ChR2 was positioned directly above the NAc for optical stimulation (Supplementary Fig. 3). Mice were then placed in behavioural testing chambers equipped with two ports: an active port, which when triggered by beam-breaks from nose-poke responses, produced an optical stimulation train to activate BLA-to-NAc fibres

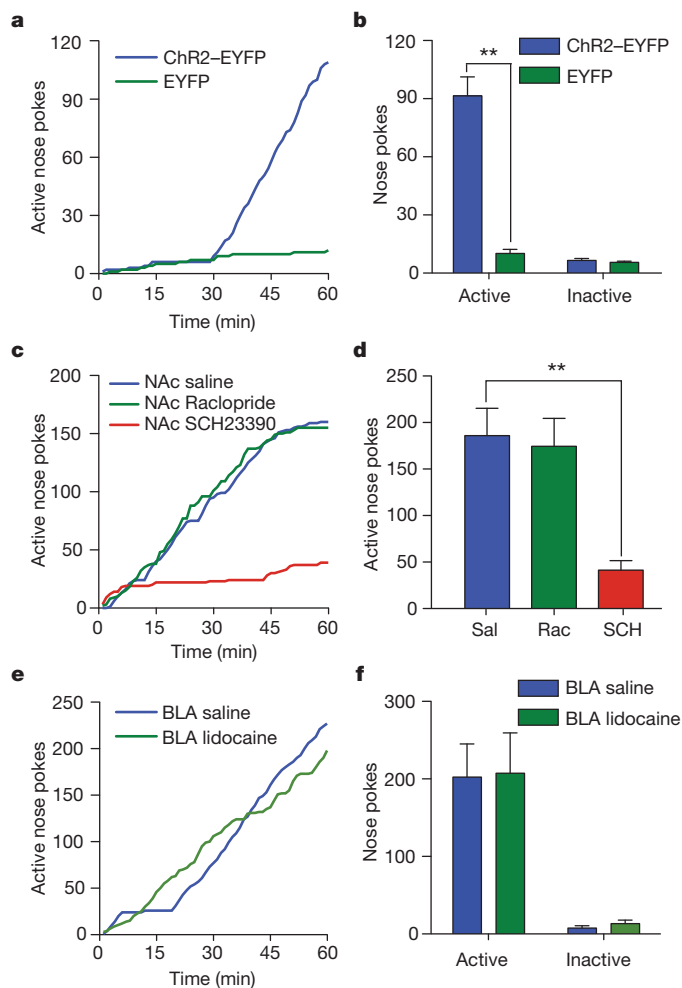


**Figure 1 | Expression of ChR2-EYFP in BLA neurons and fibres projecting to the NAc.** **a**, Coronal brain slice stained with red fluorescent Nissl stain, showing expression of ChR2-EYFP (green) after virus injection into the BLA. D, dorsal; V, ventral; M, medial; L, lateral. **b**, Example traces and average data for action potentials in current-clamped ChR2-expressing BLA neurons in response to 5-ms light pulses ( $n = 7$  cells,  $P = 0.015$ ). **c**, Brain slice showing expression of ChR2-EYFP in the NAc after virus injection into the BLA. **d**, EPSCs recorded from NAc neurons after optical stimulation of BLA-to-NAc fibres before and after bath application of CNQX ( $n = 4$  cells,  $P = 0.007$ ). All error bars for all figures correspond to the s.e.m.

<sup>1</sup>Department of Psychiatry and Department of Cell and Molecular Physiology, UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>2</sup>Ernest Gallo Clinic and Research Center, Department of Neurology, Wheeler Center for the Neurobiology of Drug Addiction, University of California San Francisco, San Francisco, California 94608, USA. <sup>3</sup>Department of Bioengineering and Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA. <sup>4</sup>Intramural Research Program, National Institute on Drug Abuse, Baltimore, Maryland 21224, USA.

selectively, and an inactive port which produced no optical stimulation. Mice expressing Chr2–EYFP in BLA-to-NAc terminals readily learned to perform nose-poke responses to earn optical stimulations in a single 60-min behavioural session, in contrast to EYFP-expressing control mice (Fig. 2a, b, Supplementary Fig. 4 and Supplementary Movie 1). Inactive nose-poke responses were not significantly different between mice expressing Chr2–EYFP and EYFP alone, indicating that optical stimulation of BLA-to-NAc fibres did not cause an increase in general responding (Fig. 2b). In contrast, direct optical activation of BLA cell bodies was highly variable in promoting self-stimulation behaviour (Supplementary Fig. 5).

To determine whether optical stimulation of BLA-to-NAc fibres reinforced nose-poke behaviour and thus increased the likelihood of additional behavioural responses, laser stimulations were withheld while active nose-poke responses were recorded in a behavioural session. Mice expressing Chr2–EYFP showed a significant decrease



**Figure 2 | In vivo optical activation of BLA-to-NAc fibres promotes self-stimulation.** **a**, Example cumulative-activity graphs of active nose pokes made in the first behavioural session to obtain optical stimulation of BLA-to-NAc fibres in a Chr2–EYFP-expressing mouse and a control (EYFP-expressing) mouse. **b**, Average numbers of nose pokes during the first optical self-stimulation session ( $n = 12$  Chr2–EYFP mice;  $n = 10$  EYFP mice; \*\*,  $P < 0.0001$ ). **c**, Example cumulative-activity graphs of nose pokes made for optical stimulation after unilateral intra-NAc microinjections of saline, raclopride or SCH23390. **d**, Average numbers of nose pokes after intra-NAc microinjections of saline (Sal), raclopride (Rac) or SCH23390 (SCH) ( $n = 19$  saline,  $n = 11$  SCH23390,  $n = 20$  raclopride; \*\*,  $P = 0.0016$ ). **e**, Example cumulative-activity graphs of active nose pokes made for optical stimulation in mice that received intra-BLA vehicle or lidocaine. **f**, Average numbers of nose pokes after intra-BLA injection of vehicle or lidocaine ( $n = 6$  intra-BLA saline group;  $n = 6$  intra-BLA lidocaine,  $P = 0.88$ ).

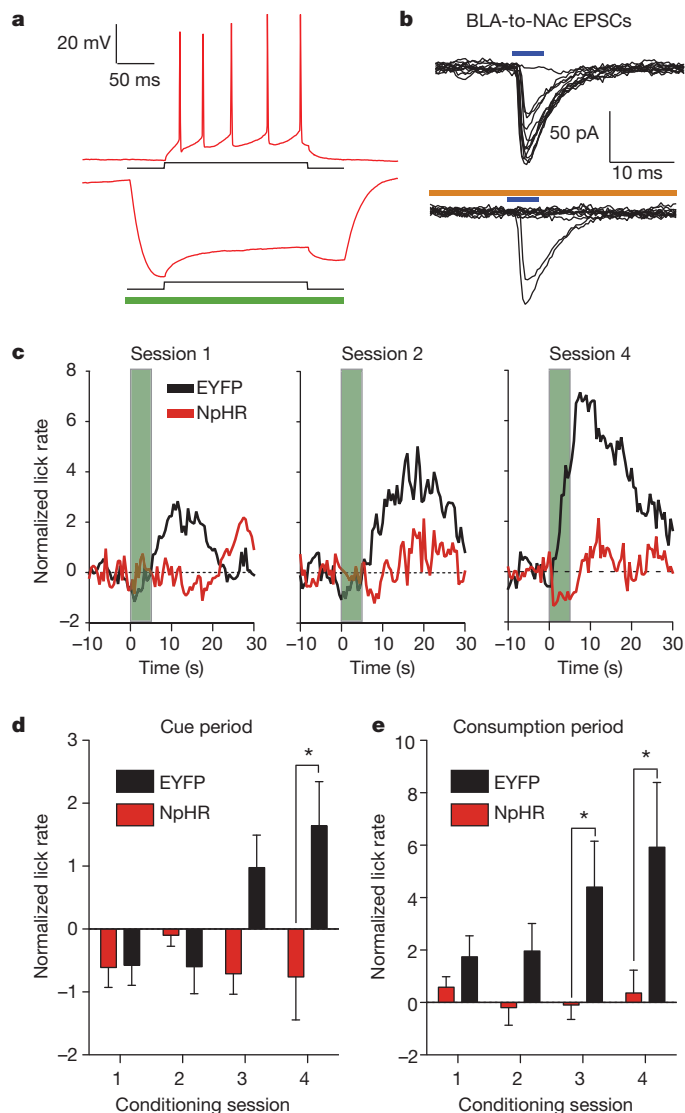
in responding when optical stimulations were withheld for 1 h (Supplementary Fig. 6). In addition, mice showed a rapid renewal of self-stimulation behaviour when optical stimuli were delivered non-contingently, and subsequent nose-poke responses were again reinforced after the 1-h extinction period.

Many forms of motivated behaviour depend on dopaminergic<sup>9–11</sup> as well as glutamatergic signalling in the NAc<sup>6,7,12</sup>. To test whether optical self-stimulation behaviour of BLA-to-NAc fibres was dependent on dopaminergic signalling, mice trained previously in the optical self-stimulation task were given microinjections into the NAc (through the same guide cannula used to introduce the optical fibre) of either vehicle, a dopamine D1-type receptor (D1R) antagonist, SCH23390, or a dopamine D2-type receptor (D2R) antagonist, raclopride, immediately before optical self-stimulation sessions. D2R antagonism (tested at two doses, Supplementary Fig. 7) had no effect, whereas D1R antagonism markedly decreased the number of active nose pokes (Fig. 2c, d). D1R antagonism did not reduce the rate of responding in the beginning of the behavioural session, nor did it affect the rate of responding within a burst of nose pokes, indicating that decreased responding during the entire session was not due to locomotor impairments induced by unilateral D1R antagonism (Supplementary Fig. 8). Notably, application of SCH23390 to NAc brain slices expressing Chr2 in BLA-to-NAc fibres markedly decreased the amplitude of all EPSCs evoked by the same optical stimulation train (60 pulses at 20 Hz) that reinforced nose-poking behaviour (Supplementary Fig. 9). These data indicate that the reinforcing properties of BLA-to-NAc stimulation require glutamate release from BLA fibres, which has postsynaptic effects on medium spiny neurons that are modulated by D1Rs.

Activation of BLA-to-NAc fibres may produce action potentials that propagate back to cell bodies in the BLA and could then activate axon collaterals that project to other brain regions. Therefore, in mice trained previously to self-stimulate, we tested whether BLA-to-NAc optical self-stimulation required neural activity in the BLA, by inactivating it with intracranial injections of lidocaine immediately before self-stimulation sessions. Ipsilateral inactivation of the BLA had no effect on acquisition (Supplementary Fig. 10) or expression of optical self-stimulation behaviour (Fig. 2e, f), demonstrating that the reinforcing properties of the optical stimulation were mediated by BLA glutamatergic fibres in the NAc or by fibre collaterals outside the BLA.

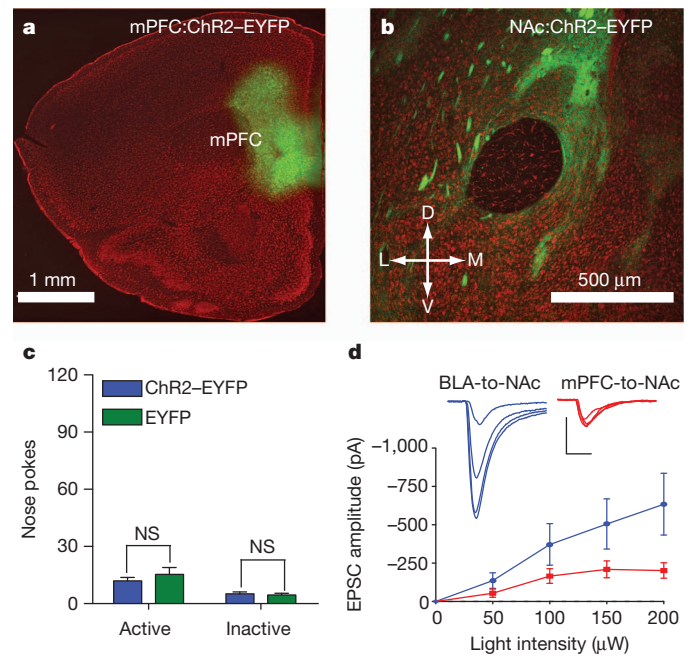
To determine whether the activity of BLA-to-NAc fibres was required for naturally occurring motivational processing, we performed pathway-specific optical inactivation experiments in a separate behavioural task in which mice were trained to drink a sucrose solution in response to a reward-predictive cue. The BLA was bilaterally injected with a virus encoding the light-gated  $\text{Cl}^-$  pump, *Natronomonas pharaonis* halorhodopsin (NpHR)<sup>13</sup> (AAV-Camk2 $\alpha$ -eNpHR3.0–EYFP; Supplementary Fig. 11). Whole-cell recordings from brain slices containing NpHR-expressing BLA neurons showed that 500-ms pulses of 532-nm light delivered to the slice resulted in prominent outward currents ( $146.2 \pm 61.4$  pA;  $n = 5$  cells) when neurons were voltage-clamped at  $-60$  mV. Current injections that reliably produced trains of action potentials were inefficient at eliciting spiking when NpHR was activated (Fig. 3a). In a subset of mice in which BLA neurons were transduced with viruses to express both Chr2 and NpHR, stimulation of BLA-to-NAc fibres via activation of Chr2 with 473-nm light resulted in light-evoked EPSCs, as predicted. However, when NpHR was simultaneously active in BLA-to-NAc fibres, Chr2 activation resulted in markedly more failed EPSCs (Fig. 3b). Thus, NpHR activation was capable of reducing evoked BLA-to-NAc EPSCs, and should therefore also reduce the endogenous activity of BLA-to-NAc fibres *in vivo*.

Mice with optical fibres implanted above the NAc and expressing NpHR in BLA-to-NAc fibres underwent four conditioning sessions consisting of 50 trials in which a 5-s tone and house-light stimulus predicted the delivery of 20  $\mu\text{l}$  of 20% sucrose. Motivated behavioural



**Figure 3 | *In vivo* optical inactivation of BLA-to-NAc fibres reduces behavioural responding for sucrose.** **a**, Injection of 100 pA current for 200 ms into NpHR-expressing neurons in the BLA results in reliable spiking of BLA neurons ( $6.6 \pm 0.9$  spikes). In all neurons, NpHR-mediated hyperpolarization completely blocked spikes due to the current injection ( $P = 0.02$ ,  $n = 3$ ). **b**, ChR2 (473 nm)-evoked EPSCs at BLA-to-NAc synapses are reduced when NpHR is activated (593.5 nm) in the same pathway. **c**, Average normalized lick rates (Z-score), time-locked to cue onset ( $t = 0-5$  s, green bar) and sucrose delivery ( $t = 5$  s), for NpHR-expressing and EYFP-expressing mice. BLA-to-NAc fibres were transiently inactivated (from  $t = -0.2$  s to  $t = 5.2$  s) in NpHR-expressing mice on each trial of each conditioning session. **d**, **e**, Data from panel **c** divided into time bins corresponding to the cue period ( $t = 0-5$  s) or the sucrose consumption period ( $t = 5-15$  s). Lick rates were significantly attenuated during the cue period (**d**) in mice receiving BLA-to-NAc inhibition ( $P = 0.013$  for treatment,  $n = 7$  mice per group). Lick rates were also significantly reduced during the sucrose consumption period (**e**) ( $P = 0.001$  for treatment,  $n = 7$  mice per group).

responding was assayed by the number of licks that each mouse made at the sucrose receptacle. On each cue–reward pairing, BLA-to-NAc fibres were transiently inactivated by delivering laser pulses bilaterally 200 ms before cue onset and terminating these pulses 200 ms after the end of the cue (laser on for 5.4 s per trial, Supplementary Fig. 12). Laser illumination was delivered in an identical fashion to control mice expressing only EYFP. Over the four conditioning sessions, control mice developed robust time-locked licking behaviour in response to the reward-predictive stimulus as well as to subsequent sucrose



**Figure 4 | *In vivo* optical activation of mPFC-to-NAc fibres does not promote self-stimulation.** **a**, Coronal brain slice stained with red fluorescent Nissl showing expression of ChR2-EYFP (green) after virus injection into the mPFC. **b**, Expression of ChR2-EYFP in fibres originating in the mPFC and innervating the NAc. **c**, Average numbers of nose pokes made by mice expressing ChR2-EYFP in mPFC-to-NAc fibres and by control (EYFP-expressing) mice ( $n = 12$  ChR2-EYFP mice;  $n = 10$  EYFP mice; NS, not significant,  $P = 0.333$ ). **d**, EPSCs recorded from NAc neurons, evoked by either mPFC-to-NAc or BLA-to-NAc optical stimulation at increasing light intensities ( $n = 7$  cells per group; effect for stimulated input,  $P = 0.003$ ).

delivery (Fig. 3c–e). In contrast, mice that received transient inhibition of BLA-to-NAc fibres during the cue–reward pairing period showed a marked attenuation of licking in response to the cue or to subsequent reward delivery (Fig. 3c–e). Transient BLA-to-NAc inactivation during cue–reward pairing also reduced the total number of licks throughout the entire session. However, when NpHR-expressing mice underwent an additional sucrose-responding session, but without laser inhibition of the BLA-to-NAc fibres, the amount of licking in the session returned to levels similar to those observed in control mice, demonstrating that the presence of NpHR alone (without optical modulation) was not sufficient to alter licking behaviour (Supplementary Fig. 13). These data show that brief, transient inhibition of BLA-to-NAc fibres can reduce motivated behavioural responding to obtain natural rewards.

In addition to the glutamatergic projection from the BLA, the NAc receives excitatory synaptic inputs from infralimbic and prelimbic regions of the medial prefrontal cortex (mPFC)<sup>14</sup> that are thought to modulate compulsive reward-seeking behaviour<sup>15,16</sup>. To determine whether activation of mPFC-to-NAc excitatory synaptic connections promotes reward-seeking behaviour similarly to BLA-to-NAc activation, mice were injected into the mPFC with ChR2-EYFP virus (Fig. 4a). This resulted in expression of ChR2-EYFP in fibres in the NAc (Fig. 4b). Mice were then tested to determine whether optical activation of mPFC-to-NAc fibres (Supplementary Fig. 14) supported self-stimulation behaviour similar to that caused by BLA-to-NAc activation. Notably, mice expressing ChR2-EYFP at mPFC-to-NAc connections showed no difference in the numbers of active or inactive nose pokes made relative to EYFP-expressing control mice (Fig. 4c). ChR2-EYFP-expressing mPFC neurons were optically excitable (Supplementary Fig. 15) and optically evoked mPFC-to-NAc EPSCs were readily detectable (Supplementary Fig. 16), demonstrating that optical activation of the mPFC-to-NAc inputs induced glutamate release, but did not support optical self-stimulation.



To determine whether quantitative differences existed in the amount of glutamate released from these two pathways, fluorescence-guided, whole-cell recordings in the NAc were performed while varying the light-stimulus intensity, in separate groups of mice selectively expressing Chr2 in either the BLA or the mPFC-to-NAc pathway. In NAc neurons that showed clear light-evoked EPSCs, BLA-to-NAc-evoked EPSCs had approximately twice the amplitude of those evoked after mPFC-to-NAc stimulation at maximal light intensities (Fig. 4d). In addition, NAc neurons typically showed excitatory postsynaptic responses to both optical stimulation of BLA inputs and electrical stimulation of cortical afferents (Supplementary Fig. 17), indicating that medium spiny neurons in the NAc receive both mPFC and BLA inputs, but that mPFC inputs release less glutamate.

These results show that selective activation of BLA, but not mPFC, glutamatergic inputs to the NAc promotes motivated behavioural responding. This is consistent with the hypothesized role of BLA inputs in facilitating responding to cues, and of mPFC inputs in suppressing inappropriate actions<sup>16</sup>. Dopamine signalling that is capable of activating D1Rs during optical self-stimulation sessions could arise from the burst-firing of dopaminergic neurons, time-locked to salient stimuli during behavioural responding<sup>17,18</sup>. Alternatively, glutamate released from BLA terminals may gate the release of dopamine from dopaminergic fibres in the NAc directly, independently of neuronal activity in the ventral tegmental area<sup>19,20</sup>. Our results show that afferent-specific glutamatergic neurotransmission from the BLA to the NAc is both necessary and sufficient to promote the expression of motivated behavioural responding.

## METHODS SUMMARY

**Opsin delivery to neural tissue.** The adeno-associated viruses AAV-Camk2 $\alpha$ -Chr2-EYFP, AAV-Camk2 $\alpha$ -EYFP and AAV-Camk2 $\alpha$ -NpHR3.0-EYFP were packaged as AAV5 by the University of North Carolina vector core facility. Virus (0.5  $\mu$ l) was stereotactically injected into the BLA or mPFC at a rate 0.1  $\mu$ l min<sup>-1</sup> via 26-gauge injector needles coupled to a 2  $\mu$ l Hamilton syringe. Mice were used for experiments about 28 d after virus injections.

**Brain-slice electrophysiology.** Opsin-expressing mice were deeply anaesthetized, decapitated and 200  $\mu$ m sections of the BLA, NAc or mPFC were prepared. Whole-cell voltage-clamp recordings were performed using a caesium methylsulphonate internal solution and current-clamp recordings were performed using a potassium gluconate internal solution. One to five ms of 473-nm, 532-nm or 593.5-nm light was delivered via a fibre-coupled laser.

**In vivo optogenetic stimulation and inhibition during behaviour.** Mice injected with opsin-encoding viral constructs were implanted with guide cannulae or chronic optical fibres directly above the NAc. Acute or chronic optical implants were connected to optical patch cables coupled to 473-nm or 532-nm lasers that were modulated by a stimulus pulse generator. The onset of laser pulses was controlled by signal pulses generated by behavioural hardware (Med Associates).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 January; accepted 10 May 2011.

Published online 29 June 2011.

1. Balleine, B. W. & Killcross, S. Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.* **29**, 272–279 (2006).
2. Maren, S. & Quirk, G. J. Neuronal signalling of fear memory. *Nature Rev. Neurosci.* **5**, 844–852 (2004).
3. LeDoux, J. The emotional brain, fear, and the amygdala. *Cell. Mol. Neurobiol.* **23**, 727–738 (2003).
4. Cador, M., Robbins, T. W. & Everitt, B. J. Involvement of the amygdala in stimulus-reward associations: Interaction with the ventral striatum. *Neuroscience* **30**, 77–86 (1989).

5. Tye, K. M., Stuber, G. D., de Ridder, B., Bonci, A. & Janak, P. H. Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. *Nature* **453**, 1253–1257 (2008).
6. Ambroggi, F., Ishikawa, A., Fields, H. L. & Nicola, S. M. Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* **59**, 648–661 (2008).
7. Di Ciano, P. & Everitt, B. J. Direct interactions between the basolateral amygdala and nucleus accumbens core underlie cocaine-seeking behavior by rats. *J. Neurosci.* **24**, 7167–7173 (2004).
8. Zhang, F., Wang, L. P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3**, 785–792 (2006).
9. Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M. & Carelli, R. M. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003).
10. Stuber, G. D. *et al.* Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science* **321**, 1690–1692 (2008).
11. Tsai, H. C. *et al.* Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**, 1080–1084 (2009).
12. Di Ciano, P., Cardinal, R. N., Cowell, R. A., Little, S. J. & Everitt, B. J. Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *J. Neurosci.* **21**, 9471–9477 (2001).
13. Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165 (2010).
14. Wright, C. I. & Groenewegen, H. J. Patterns of convergence and segregation in the medial nucleus accumbens of the rat: relationships of prefrontal cortical, midline thalamic, and basal amygdaloid afferents. *J. Comp. Neurol.* **361**, 383–403 (1995).
15. McFarland, K., Lapish, C. C. & Kalivas, P. W. Prefrontal glutamate release into the core of the nucleus accumbens mediates cocaine induced reinstatement of drug-seeking behavior. *J. Neurosci.* **23**, 3531–3537 (2003).
16. Kalivas, P. W., Volkow, N. & Seamans, J. Unmanageable motivation in addiction: A pathology in prefrontal-accumbens glutamate transmission. *Neuron* **45**, 647–650 (2005).
17. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
18. Bromberg-Martin, E. S. & Hikosaka, O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* **63**, 119–126 (2009).
19. Floresco, S. B., Yang, C. R., Phillips, A. G. & Blaha, C. D. Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. *Eur. J. Neurosci.* **10**, 1241–1251 (1998).
20. Jones, J. L. *et al.* Basolateral amygdala modulates terminal dopamine release in the nucleus accumbens and conditioned responding. *Biol. Psychiatry* **67**, 737–744 (2010).
21. Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870 (2006).
22. Shabel, S. J. & Janak, P. H. Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proc. Natl Acad. Sci. USA* **106**, 15031–15036 (2009).
23. Tye, K. M. *et al.* Amygdala circuitry mediating reversible and bidirectional control of anxiety. *Nature* **471**, 358–362 (2011).
24. Shiflett, M. W. & Balleine, B. W. At the limbic-motor interface: disconnection of basolateral amygdala from nucleus accumbens core and shell reveals dissociable components of incentive motivation. *Eur. J. Neurosci.* **32**, 1735–1743 (2010).
25. Setlow, B., Holland, P. C. & Gallagher, M. Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive pavlovian second-order conditioned responses. *Behav. Neurosci.* **116**, 267–275 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Phillips, V. Kharazia, A. Adamantidis and H.-C. Tsai for assistance and advice. We also thank V. Gukasyan and the UNC Neuroscience Center microscopy core facility. This study was supported by funds from NARSAD, ABMRF, The Foundation of Hope, and NIDA (DA029325), by startup funds provided by the Psychiatry Department at UNC Chapel Hill (G.D.S.) and by the State of California through the University of California at San Francisco (A.B.). D.R.S. was supported by F32AA018610.

**Author Contributions** G.D.S. and A.B. designed, discussed and planned all experiments. G.D.S., D.R.S., A.M.S., W.A.v.L., J.E.H., S.C., K.M.T. and K.A.K. performed experiments. G.D.S., D.R.S., A.M.S. and W.A.v.L. analysed data. F.Z. and K.D. provided resources and training to G.D.S. G.D.S. and A.B. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.D.S. ([gstuber@med.unc.edu](mailto:gstuber@med.unc.edu)).

## METHODS

**Experimental subjects and stereotaxic surgery.** Adult (25–30 g) male C57BL/6J mice (Jackson Laboratory) were group-housed until surgery. Mice were maintained on a 12 h:12 h light:dark cycle (lights on at 7:00). After the animals were acclimatized to the animal facility for ~1 week, they were anaesthetized with 150 mg kg<sup>-1</sup> ketamine and 50 mg kg<sup>-1</sup> xylazine and placed in a stereotaxic frame (Kopf Instruments). Microinjection needles were then inserted bilaterally directly above the BLA (coordinates from Bregma: -1.6 AP,  $\pm$ 3.1 ML, -4.9 DV). Microinjections were performed using custom-made injection needles (26-gauge) connected to a 2- $\mu$ l Hamilton syringe. Each BLA was injected with 0.3–0.5  $\mu$ l of purified and concentrated AAV (~10<sup>12</sup> infectious units ml<sup>-1</sup>) encoding ChR2-EYFP, NpHR3.0-EYFP or EYFP alone under the control of the *Camk2 $\alpha$*  promoter. Injections occurred over 10 min followed by an additional 10 min to allow diffusion of viral particles away from the injection site. For optical self-stimulation experiments, mice were first injected unilaterally into the BLA with virus and then a guide cannula was implanted directly over the ipsilateral NAc (+1.3 AP,  $\pm$ 1.0 ML, -4.0 DV) to allow insertion of the fibre-optic cable during the experiment. The fibre was secured to the skull using Geristore (<http://www.denmat.com>) dental cement. Mice were then returned to their home cage. Body weight and signs of illness were monitored until recovery from surgery (approximately 2 weeks). All procedures were conducted in accordance with the guide for the care and use of laboratory animals, as adopted by the NIH, and with approval of the UNC and UCSF institutional animal care and use committees.

**Construct and AAV preparation.** DNA plasmids encoding pAAV-Camk2 $\alpha$ -ChR2-EYFP (H134R), pAAV-Camk2 $\alpha$ -NpHR3.0-EYFP or pAAV-Camk2 $\alpha$ -EYFP were obtained from the laboratory of K. Deisseroth (see <http://www.optogenetics.org> for additional details). Plasmid DNA was amplified, purified and collected using a standard plasmid maxiprep kit (Qiagen). After plasmid purification, restriction digest and sequencing to confirm DNA fidelity, purified recombinant AAV vectors were serotyped with AAV5 coat proteins and packaged by the UNC vector core facilities using calcium phosphate precipitation methods. The final viral concentration was  $1-2 \times 10^{12}$  viral particles ml<sup>-1</sup>.

**Slice preparation for patch-clamp electrophysiology.** Mice were anaesthetized with pentobarbital and perfused transcardially with modified artificial cerebrospinal fluid containing 225 mM sucrose, 119 mM NaCl, 2.5 mM KCl, 1.0 mM NaH<sub>2</sub>PO<sub>4</sub>, 4.9 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>, 26.2 mM NaHCO<sub>3</sub> and 1.25 mM glucose. The brain was removed rapidly from the skull and placed in the same solution used for perfusion, at ~0 °C. Coronal sections of the NAc or BLA (200  $\mu$ m) were then cut on a vibratome (VT-1200, Leica Microsystems). Slices were placed in a holding chamber and allowed to recover for at least 30 min before being placed in the recording chamber and superfused with bicarbonate-buffered solution saturated with 95% O<sub>2</sub> and 5% CO<sub>2</sub> and containing 119 mM NaCl, 2.5 mM KCl, 1.0 mM NaH<sub>2</sub>PO<sub>4</sub>, 1.3 mM MgCl<sub>2</sub>, 2.5 mM CaCl<sub>2</sub>, 26.2 mM NaHCO<sub>3</sub> and 11 mM glucose (at ~32 °C).

**Patch-clamp electrophysiology.** Cells were visualized using infrared differential interference contrast and fluorescence microscopy. Whole-cell voltage-clamp or current-clamp recordings of BLA and NAc neurons were made using an Axopatch 200A or B amplifier. Patch electrodes (3.0–5.0 M $\Omega$ ) were backfilled with internal solution containing 130 mM KOH, 105 mM methanesulphonic acid, 17 mM hydrochloric acid, 20 mM HEPES, 0.2 mM EGTA, 2.8 mM NaCl, 2.5 mg ml<sup>-1</sup> MgATP and 0.25 mg ml<sup>-1</sup> GTP (pH 7.35, 270–285 mOsm). Series resistance (15–25 M $\Omega$ ) and/or input resistance were monitored online with a 4 mV hyperpolarizing step (50 ms) given between stimulation sweeps. All data were filtered at 2 kHz, digitized and collected using pClamp10 software (Molecular Devices). For current-clamp experiments to characterize cell firing, ten pulses at frequencies of 1, 5, 10 and 20 Hz, respectively, were tested to determine spike fidelity (the percentage of light pulses that lead to action potentials). For optical stimulation of EPSCs, stimulation (pulses of 1–2 mW, 473-nm light delivery via a 200- $\mu$ m optical fibre coupled to a solid-state laser) was used to evoke presynaptic glutamate release from BLA projections to the NAc. NAc medium spiny neurons were voltage-clamped at -70 mV. For pharmacological characterization of glutamate currents, light-evoked EPSCs were recorded for 10 min, followed by bath application of 10  $\mu$ M CNQX for an additional 10 min. Ten to twelve sweeps before and after drug application were averaged and peak EPSC amplitudes were then measured. For EPSC pulse-train experiments, input-specific currents were evoked by 60 optical pulses (20 Hz stimulation, 5 ms pulse duration). This was repeated 12 times at 0.1 Hz. SCH23390 (4  $\mu$ M) or vehicle was then bath-applied for 10 min and the stimulus train was repeated. The average EPSC train from the six sweeps immediately before drug application was then compared with the train for the six sweeps immediately after drug application.

**In vivo optrode recording.** Approximately 21–28 d after bilateral injection of AAV-Camk2 $\alpha$ -ChR2-EYFP into the BLA, mice were deeply anaesthetized with ketamine and xylazine and placed in a stereotaxic frame equipped with a temperature

controller to regulate body temperature. The skull was then removed directly above the NAc. Parylene-coated tungsten electrodes (1 M $\Omega$ ), attached with epoxy resin to an optical fibre of 200  $\mu$ m core diameter and 0.37 numerical aperture coupled to a 473-nm laser, were then lowered into the NAc to record unit activity of postsynaptic medium spiny neurons after trains of light pulses were used to evoke BLA-to-NAc-specific glutamate release. Ten pulses of light (10–20 mW, 5 ms) at frequencies of 1, 5, 10 and 20 Hz, respectively, were used to determine spike fidelity *in vivo*, analogous to the experiment performed during whole-cell recording. Unit activity was amplified with an extracellular amplifier (A-M systems), band-pass filtered at 300 Hz low/5 kHz and digitized using pClamp10 software.

**Freely moving optical self-stimulation.** At 21–28 d after injection of pAAV-Camk2 $\alpha$ -ChR2-EYFP or control virus into the BLA, mice with cannulae placed above the NAc were prepared for nose-poke training. Mice were mildly food-restricted to 4 g of food per day to stabilize body weight and facilitate behavioural responding. Body weight was monitored throughout the experiment and did not fall below ~90% of their free-feeding weight. Immediately before placing mice in the operant chambers, stylets were removed from the cannulae and a flat-cut 125- $\mu$ m-diameter fibre-optic cable, coupled to a solid-state 473-nm laser outside the operant chamber, was inserted through the guide cannula and placed directly above the NAc. Immediately before insertion through the guide cannula, light output through the optical fibres was adjusted to 10–20 mW. The optical fibre was then secured into place via a custom-made locking mechanism to ensure that no movement of the fibre occurred during the experiment. Mice were then placed in standard Med-Associates operant chambers equipped with an active and inactive nose-poke operandum directly below two cue lights. The chambers were also equipped with house lights, audio stimulus generators and video cameras coupled to DVD recorders. A 1-h optical self-stimulation session began with the onset of the cue light above the active nose-poke operandum. Each active nose poke performed by the animal resulted in an optical stimulation of BLA-to-NAc fibres (60 pulses, 20 Hz, 5 ms pulse duration). Both active and inactive nose-poke timestamp data were recorded using Med-PC software and analysed using Neuroexplorer and Microsoft Excel software.

**NAc microinjections before optical self-stimulation.** Stylets were removed from guide cannulae and a 26-gauge injector needle connected to a 1- $\mu$ l Hamilton syringe was inserted. All microinjections were delivered in 0.3  $\mu$ l sterile saline at a rate of 0.1  $\mu$ l min<sup>-1</sup>. Injector needles remained in place for an additional 2 min before being removed and replaced immediately with stylets or optical fibres for self-stimulation sessions. Doses of drugs used for microinjections were: 600 ng in 0.3  $\mu$ l for SCH23390; 100 ng in 0.3  $\mu$ l and 3  $\mu$ g in 0.3  $\mu$ l for raclopride; and 10  $\mu$ g in 0.3  $\mu$ l for lidocaine.

**Implantable optical fibres for NpHR inhibition during behaviour.** For these experiments, mice were bilaterally injected into the BLA with virus encoding NpHR3.0-EYFP or EYFP, as described above. Mice were also implanted with bilateral optical fibres targeted directly above each NAc. Optical fibres were constructed in-house by interfacing a 7–10-mm piece of 200- $\mu$ m, 0.37-numerical-aperture optical fibre with a 1.25-mm zirconia ferrule (fibre extending 5 mm beyond the end of the ferrule). Fibres were attached with epoxy resin into the ferrules, then cut and polished. After construction, all fibres were calibrated to determine a percentage of light transmission at the fibre tip that would interface with the brain. Before bilateral implantation, fibres were matched to each other so that each fibre would output an equal amount of light (to within 10%). This was done to ensure that an equal amount of light was delivered to each hemisphere. After surgery, protective plastic caps were placed on the implanted optical fibres to protect them from dust and debris.

Four to five weeks after implantation surgery and 3 d before the experiment, mice were connected to 'dummy' optical-patch cables each day for 30–60 min to habituate them to the tethering procedure in their home cage. On experiment days, protective caps were removed from the implanted fibres. Fibres were then connected to custom-made optical-patch cables (62.5  $\mu$ m core diameter) that were covered with furcation tubing to protect the cables and prevent light from the laser from illuminating the operant chamber. Bilateral fibres were connected to a fibre splitter (50:50 split ratio) that interfaced with a fibre-coupled 532-nm DPSS laser (200 mW). On the basis of the calibration factor of each pair of fibres, light intensity was set to 10 mW illumination at each fibre tip in the brain.

**Optical inhibition of BLA-to-NAc fibres during sucrose responding.** Mice with optical fibres implanted above the NAc, and expressing either NpHR3.0-EYFP or EYFP in BLA-to-NAc fibres, were trained to drink sucrose in response to an environmental stimulus that predicted sucrose delivery. The start of the session was signalled by the onset of white noise in the operant chamber. Each session consisted of 50 cue-reward pairings with a random inter-trial interval of 120 s. During each trial, a digital pulse was sent from the behavioural hardware to engage the laser 200 ms before the onset of a 5-s reward-predictive stimulus (tone/house-light compound stimulus). Delivery of 20  $\mu$ l of 20% sucrose to a receptacle

occurred immediately after the termination of the reward-predictive cue, and the laser pulse was terminated 200 ms after the cue ended. The laser pulse was started and extended for 200 ms before and after the cue on the basis of *in vitro* experiments in which we observed that activation of NpHR led to maximal inhibition 200 ms after the start of the laser pulse. Cue presentation, reward delivery, lick and laser time-stamps were stored as separate data arrays and analysed offline with Microsoft Excel and Neuroexplorer.

Time-locked licking behaviour was quantified for all mice. Mice that did not make at least 200 licks on at least one of the four conditioning sessions were excluded from analysis. This resulted in the removal of two NpHR and two EYFP mice from analysis. Time-locked lick histograms with 0.5-s time bins were then constructed from  $-10$  s to  $30$  s, time-locked to the cue onset ( $t = 0$ ). Lick rates were normalized to baseline periods using a Z-score procedure ( $z = (x - \mu)/\sigma$ ) with  $\mu$  being the average lick rate and  $\sigma$ , the standard deviation in the  $10$  s preceding the cue onset.

**Data analysis.** Statistical significance was assessed using *t*-tests or analysis of variance (ANOVA), followed by post-hoc tests when applicable, using  $\alpha = 0.05$ . Data were analysed using Microsoft Excel with the Statplus plugin and Prism (GraphPad Software).

**Virus expression and histology.** After behavioural experiments, mice were deeply anaesthetized with pentobarbital and perfused transcardially with PBS followed by 4% paraformaldehyde dissolved in PBS. Brains were removed carefully and fixed in 4% paraformaldehyde for an additional 24–48 h. Brains were transferred to 30% sucrose for 48–72 h before slicing  $50\ \mu\text{m}$  sections of the BLA or NAc on a freezing-stage microtome or cryostat. Slices were then washed three times in PBS for 5 min. Slices were then stained for 1 h with 2% Neurotrace fluorescent Nissl stain (Invitrogen; excitation 530 nm, emission 615 nm) diluted in PBS with 0.1% Triton X-100. Slices were then washed and mounted on gelatin-coated slides, treated with fluorescent-mounting media and mounted. Expression of ChR2-EYFP, NpHR3.0-EYFP or EYFP was then examined for all mice using either a Nikon inverted fluorescent microscope with a  $\times 4$ ,  $\times 10$  or  $\times 20$  objective or a Zeiss laser-scanning confocal microscope at  $\times 25$  and  $\times 63$ . After injection of virus into

the BLA, robust expression of ChR2-EYFP was observed in BLA projection targets including the NAc, mPFC, hippocampus, insular cortex and to a lesser extent, the dorsal medial striatum. Mice showing no EYFP expression in the NAc owing to faulty microinjections, and mice showing cannula or fibre placements outside the NAc, were excluded from analysis.

**Reconstruction of optical stimulation or inhibition sites in the NAc.** To determine optical stimulation sites in experiments in which guide cannulae were used to introduce optical fibres into brain tissue (BLA-to-NAc and mPFC-to-NAc optical self-stimulation experiments, see Supplementary Figs 3 and 14 for the location of optical stimulation sites), fixed and stained coronal brain sections (see above) containing the NAc and cannula tracks were examined on an upright conventional fluorescent microscope. Cannula tracks were located in the slices and optical stimulation sites were determined by locating the site 1 mm ventral to the end of the cannula tip. A 1-mm distance was used in these experiments because the optical fibres extended 0.5 mm beyond the end of the cannula (each fibre was cut to this length before insertion). On the basis of the light output from these optical fibres ( $477\ \text{mW mm}^{-1}$  at the tip), and calculating intensity by taking into account geometric loss and scattering through tissue<sup>1</sup>, loss at 0.5 mm beyond the fibre tip led to an estimated 2.6% transmission, or  $12.4\ \text{mW mm}^{-1}$  at this distance. At 1 mm from the tip of the optical fibre, estimated transmission dropped to 0.56% or  $2.67\ \text{mW mm}^{-1}$ , which approximates the minimum intensity required to activate opsin proteins ( $1\ \text{mW mm}^{-1}$ ). For NpHR-mediated inhibition experiments, optical inhibition sites (Supplementary Fig. 11) were determined in a similar fashion, with 0.5 mm used as the distance from the fibre tip to the diagrammed inhibition sites because no guide cannula was present. This distance represents the centre location where optical stimulation or inhibition occurs (0.5 mm above and below). All calculations were performed using equations and constants listed in ref. 21.

26. Aravanis, A. M. *et al.* An optical neural interface: *in vivo* control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* **4**, S143–S156 (2007).



# Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis

Sacri R. Ferrón<sup>1</sup>, Marika Charalambous<sup>1\*</sup>, Elizabeth Radford<sup>1\*</sup>, Kirsten McEwen<sup>1</sup>, Hendrik Wildner<sup>2</sup>, Eleanor Hind<sup>1</sup>, Jose Manuel Morante-Redolat<sup>3</sup>, Jorge Laborda<sup>4</sup>, Francois Guillemot<sup>2</sup>, Steven R. Bauer<sup>5</sup>, Isabel Fariñas<sup>3</sup> & Anne C. Ferguson-Smith<sup>1</sup>

**The gene for the atypical NOTCH ligand delta-like homologue 1 (*Dlk1*) encodes membrane-bound and secreted isoforms that function in several developmental processes *in vitro* and *in vivo*. *Dlk1*, a member of a cluster of imprinted genes, is expressed from the paternally inherited chromosome<sup>1,2</sup>. Here we show that mice that are deficient in *Dlk1* have defects in postnatal neurogenesis in the subventricular zone: a developmental continuum that results in depletion of mature neurons in the olfactory bulb. We show that DLK1 is secreted by niche astrocytes, whereas its membrane-bound isoform is present in neural stem cells (NSCs) and is required for the inductive effect of secreted DLK1 on self-renewal. Notably, we find that there is a requirement for *Dlk1* to be expressed from both maternally and paternally inherited chromosomes. Selective absence of *Dlk1* imprinting in both NSCs and niche astrocytes is associated with postnatal acquisition of DNA methylation at the germ-line-derived imprinting control region. The results emphasize molecular relationships between NSCs and the niche astrocyte cells of the microenvironment, identifying a signalling system encoded by a single gene that functions coordinately in both cell types. The modulation of genomic imprinting in a stem-cell environment adds a new level of epigenetic regulation to the establishment and maintenance of the niche, raising wider questions about the adaptability, function and evolution of imprinting in specific developmental contexts.**

The mammalian adult brain is generally postmitotic, but reservoirs of NSCs with features of astroglial cells exist in the hippocampus and subventricular zone (SVZ), supporting lifelong neurogenesis<sup>3,4</sup>. The SVZ is a very active germinal niche in which production of neurons occurs via a transit-amplifying progenitor population<sup>4</sup>, giving rise to migrating neuroblasts that integrate into the circuitry of the olfactory bulb<sup>5</sup>. The specialized microenvironment containing niche astrocytes regulates long-term maintenance of NSCs and ensures continual neurogenesis<sup>3,4,6</sup>. An emerging hypothesis is that NSCs and niche astrocytes are established in the postnatal radial glia/astrocytic lineage<sup>7</sup>; however, the potential lineage relationships and cell–cell interactions between them are not completely understood<sup>6–8</sup>.

*Dlk1* encodes a transmembrane protein belonging to the Notch/Delta/Serrate family of signalling molecules, which have key roles in differentiation<sup>9–11</sup>. *Dlk1* is widely expressed during embryonic development<sup>12–14</sup> and is dosage-sensitive, with overexpression causing phenotypes ranging from prenatal lethality to defects in postnatal energy homeostasis<sup>15</sup>. Few tissues retain *Dlk1* expression postnatally and deletion experiments have demonstrated *in vivo* functions for DLK1 in adipogenesis and haematopoiesis<sup>10,11,16</sup>. We detected DLK1 in neurogenic areas of the prenatal telencephalon and in the postnatal SVZ, with a peak at postnatal day 7 (P7). In the adult brain, *Dlk1* expression is mainly restricted to neurons of several areas, including

the ventral tegmental area<sup>14</sup>, the septum and the ventral striatum, but the protein is still detected in specific cell types of the mature SVZ, in particular NSCs (GFAP<sup>+</sup>SOX2<sup>+</sup>NESTIN<sup>+</sup>) and niche astrocytes (GFAP<sup>+</sup>SOX2<sup>+</sup>S100β<sup>−</sup>). Differentiated parenchymal astrocytes (GFAP<sup>+</sup>S100β<sup>+</sup>), βIII-tubulin<sup>+</sup> neuroblasts and IB4<sup>+</sup> ependymocytes do not contain DLK1 (Supplementary Fig. 1a–j).

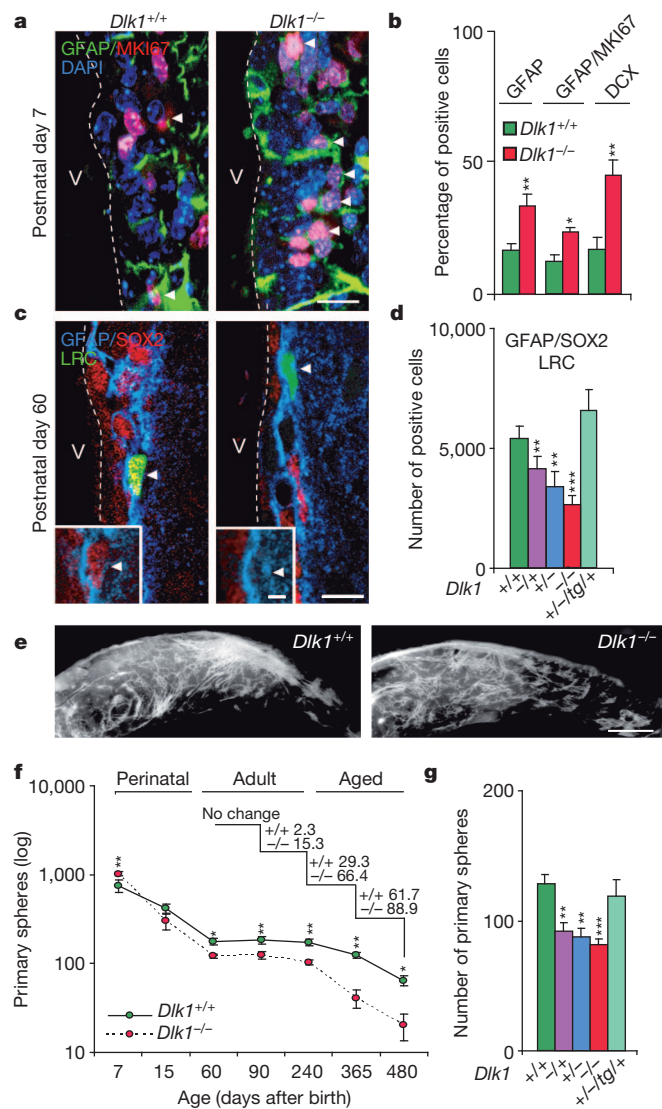
To test for potential roles for DLK1 in neurogenesis, we analysed brain germinal regions in embryos and postnatal mice with a targeted mutation in the *Dlk1* gene<sup>11</sup>. Although we observed expression of DLK1 both in progenitor cells and in differentiating neurons in ganglionic eminences at embryonic day (E)12.5 and E14.5, we did not observe any differences between *Dlk1*-wild-type and mutant embryos in progenitor-cell activity or in neurogenesis in the mantle (Supplementary Fig. 2a–c), indicating that DLK1 is dispensable during embryonic neurogenesis. In contrast, we observed increased activity of NSCs in the developing SVZ at P7, indicated by higher numbers of GFAP<sup>+</sup>MKI67<sup>+</sup> cells. This resulted in increased numbers of double-cortin (DCX)<sup>+</sup> neuroblasts (Fig. 1a, b and Supplementary Fig. 3a), indicating a failure to maintain the slower-dividing stem-cell pool in the postnatal SVZ.

Postnatal day 7 is a transition time point between development of the embryonic germinal layer and the mature SVZ<sup>17</sup>, so we next evaluated adult mice. Wild-type and *Dlk1*-mutant mice at P60 were injected with 5-bromo-2-deoxyuridine (BrdU) and killed one month later to assess BrdU-label-retaining cells (LRCs). LRCs mark relatively quiescent NSCs and cells that abandon the cell cycle shortly after labelling, such as terminally differentiated cells in the SVZ and newly-formed olfactory-bulb neurons<sup>18</sup>. The numbers of BrdU<sup>+</sup>GFAP<sup>+</sup>SOX2<sup>+</sup> or GFAP<sup>+</sup>NESTIN<sup>+</sup> cells were significantly reduced in *Dlk1*<sup>−/−</sup> mice (Fig. 1c, d and Supplementary Fig. 3b). However, the percentage of GFAP<sup>+</sup>MKI67<sup>+</sup> cells was similar in both genotypes (Supplementary Fig. 3d, f), indicating a change in NSC number but not in their cycling parameters. Fewer NSCs resulted in a smaller ASCL1 (also known as MASH1)<sup>+</sup> transit-amplifying progenitor population and fewer DCX<sup>+</sup> neuroblasts (Supplementary Fig. 3c, e, f), resulting in a less densely populated rostral migratory stream (Fig. 1e). Moreover, the numbers of postmitotic calretinin (CALB2)<sup>+</sup> and tyrosine hydroxylase (TH)<sup>+</sup> newly formed BrdU<sup>+</sup> neurons in the granular and periglomerular layers of the mutant olfactory bulb were significantly reduced (Supplementary Fig. 3g).

It has been demonstrated that disruption of quiescence at early postnatal periods leads to loss of stem-cell potential and depletion of the NSC pool later in life<sup>19,20</sup>. To test whether DLK1 was indeed required for NSC maintenance, we evaluated the size of the stem-cell pool over time by determining the yield of primary neurospheres at different ages. A transiently higher yield of primary neurospheres from P7 SVZ mutant tissue was followed by an increase in their progressive

<sup>1</sup>Department of Physiology, Development & Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK. <sup>2</sup>Department of Molecular Neurobiology, National Institute for Medical Research, Medical Research Council, London NW7 1AA, UK. <sup>3</sup>Departamento de Biología Celular, Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas, Universidad de Valencia, 46100 Burjassot, Spain. <sup>4</sup>Department of Inorganic and Organic Chemistry and Biochemistry, Medical School, Regional Center for Biomedical Research, University of Castilla-La Mancha, Avenida de Almansa 14, 02006 Albacete, Spain. <sup>5</sup>Cellular and Tissue Therapies Branch, Division of Cellular and Gene Therapies, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892, USA.

\*These authors contributed equally to this work.



**Figure 1 | DLK1 regulates postnatal neurogenesis.** **a**, Immunohistochemistry for GFAP (green) and MKI67 (red) in the P7 SVZ of wild-type and *Dlk1*<sup>-/-</sup> mice. DNA is stained with 4',6-diamidino-2-phenylindole (DAPI, blue) and arrowheads indicate positive cells. V, ventricle. **b**, Percentages of cell types in the P7 SVZ from wild-type and *Dlk1*<sup>-/-</sup> mice. **c**, Immunohistochemistry for GFAP (blue), SOX2 (red) and BrdU (LRCs, green) in the adult (P60) SVZ of wild-type and *Dlk1*<sup>-/-</sup> mice. **d**, Numbers of GFAP<sup>+</sup>SOX2<sup>+</sup>BrdU-label-retaining cells (LRCs) in the SVZ of mice with deletions of maternally inherited *Dlk1* (*Dlk1*<sup>-/+</sup>), paternally inherited *Dlk1* (*Dlk1*<sup>-/-</sup>) or both alleles (*Dlk1*<sup>-/-</sup>). The light-green bar represents paternal-transmission-mutant mice in a *Dlk1* transgenic background (*Dlk1*<sup>+/+</sup>; *Dlk1*<sup>tg/+</sup>). **e**, Whole-mounts for DCX<sup>+</sup> migrating neuroblasts in the SVZ. **f**, Numbers of primary spheres from wild-type and *Dlk1*-mutant SVZs at different developmental stages. **g**, Numbers of primary spheres from the adult SVZ of different *Dlk1* mutants. \*, *P* < 0.05; \*\*, *P* < 0.01; \*\*\*, *P* < 0.001. All error bars show s.e.m. of five experiments (*n* = 4 cultures per genotype). Scale bars: **a**, **c**, 20  $\mu$ m (inset, 10  $\mu$ m); **e**, 200  $\mu$ m.

decline as the animals aged (Fig. 1f), indicating that postnatal expression of DLK1 is required for life-long maintenance of the NSC pool.

To evaluate dosage effects of the *Dlk1* gene, mice carrying the mutation on the maternally inherited (*Dlk1*<sup>-/+</sup>) or the paternally inherited (*Dlk1*<sup>-/-</sup>) allele were analysed. Notably, the numbers of GFAP<sup>+</sup>SOX2<sup>+</sup>BrdU<sup>+</sup> LRCs and of newly generated olfactory bulb neurons were reduced in both *Dlk1* heterozygotes (Fig. 1d and Supplementary Fig. 3g). Consistent with the defects observed *in vivo*, the SVZ from *Dlk1*<sup>-/+</sup>, *Dlk1*<sup>-/+</sup> and *Dlk1*<sup>-/-</sup> mice all yielded markedly fewer primary neurospheres than wild-type tissue, although brain size

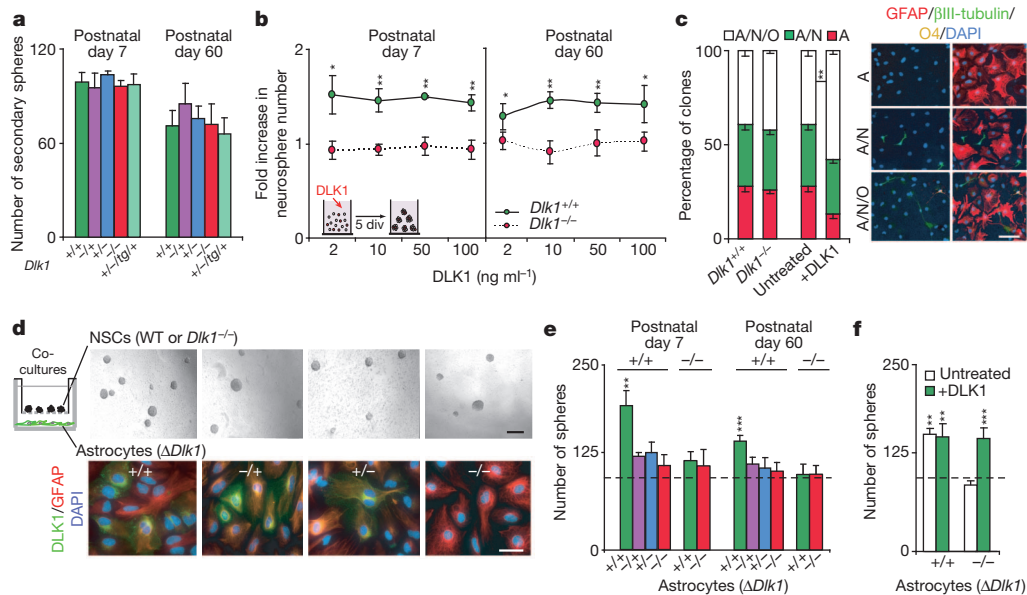
was normal (Fig. 1g and Supplementary Fig. 3h, i). We confirmed that these phenotypes were specifically due to a reduction in the available levels of DLK1 by generating paternal-transmission *Dlk1* mutants that were also hemizygous for a *Dlk1*-expressing transgene (*Dlk1*<sup>+/+</sup>; *Dlk1*<sup>tg/+</sup>)<sup>15</sup>. The number of GFAP<sup>+</sup>SOX2<sup>+</sup>BrdU<sup>+</sup> LRCs within *in vivo* and *in vitro* neurospheres obtained from double-mutant SVZ tissue was not significantly different from the number obtained from wild-type littermates, indicating that rescue of the mutant had occurred (Fig. 1d, g).

Despite being reduced in number, mutant neurospheres displayed normal clonogenic capacity upon passage (Fig. 2a), indicating that DLK1 acts as a postnatal niche-secreted factor *in vivo*. Treatment of P7 and P60 cultures with recombinant mouse DLK1 resulted in 40–50% more neurospheres (Fig. 2b), and markedly higher numbers of secondary spheres were formed from primary neurospheres that had been grown (pre-treated) in DLK1-supplemented medium (Supplementary Fig. 4a). These increases were not due to DLK1 promoting NSC survival (Supplementary Fig. 4b), indicating that the addition of DLK1 specifically increased self-renewing symmetrical divisions. Moreover, multipotentiality in clonal differentiation assays was increased by exogenous DLK1 (Fig. 2c), further supporting a role for DLK1 as a niche factor. To evaluate this further, we co-cultured astrocytes acutely isolated from the SVZ of wild-type, *Dlk1*<sup>-/+</sup>, *Dlk1*<sup>+/+</sup> and *Dlk1*<sup>-/-</sup> P7 and P60 mice<sup>8</sup> with wild-type NSCs, using transwell inserts (Fig. 2d). Wild-type niche astrocytes induced a marked increase in neurosphere formation which was abrogated when niche astrocytes were derived from *Dlk1* mutants (Fig. 2d, e). The reduction in neurosphere number in medium conditioned with *Dlk1*<sup>-/-</sup> astrocytes was rescued by the exogenous addition of recombinant DLK1 (Fig. 2f): an indication that DLK1 secreted by SVZ niche astrocytes regulates NSC self-renewal, probably in combination with other niche factors.

Alternatively spliced transcripts of *Dlk1* (Fig. 3a), encoding protein isoforms that are either membrane-tethered or proteolytically cleaved and secreted, have been described<sup>21</sup>. The secreted isoforms DLK1A and DLK1B contain a juxtamembrane motif for cleavage by extracellular proteases that is absent from membrane-bound DLK1C and DLK1D (Fig. 3a). Secreted isoforms represent the predominant type in acutely isolated niche astrocytes. Membrane-bound isoforms are preferentially expressed by P7 and P60 NSCs (Fig. 3b). Notably, exogenously added DLK1 or co-culture with niche astrocytes did not increase neurosphere formation in *Dlk1*<sup>-/-</sup> NSC cultures (Fig. 2b, e), indicating that the membrane-tethered form of DLK1 in NSCs contributes to the response to soluble DLK1.

To confirm whether membrane-bound DLK1 is required for the response to soluble DLK1, GFP-tagged vectors expressing membrane-bound (MB-DLK1) or secreted (S-DLK1) isoforms were nucleofected into *Dlk1*<sup>+/+</sup> and *Dlk1*<sup>-/-</sup> NSCs, and neurosphere formation in response to exogenous DLK1 was determined. *Dlk1*<sup>-/-</sup> NSCs regained the response to exogenous DLK1 only when expressing the membrane-bound isoform (Fig. 3c and Supplementary Fig. 4c, d). Furthermore, expression of S-DLK1 but not MB-DLK1 in *Dlk1*<sup>-/-</sup> niche astrocytes rescued the wild-type NSC response in co-cultures (Fig. 3c), indicating that membrane-bound DLK1 in NSCs is stimulated by secreted DLK1. Notably, increased expression of *Dlk1* above wild-type levels did not influence the response of NSCs (Fig. 3c and Supplementary Fig. 4e). It is worth noting that responses elicited by Jagged 1 (ref. 22) and SERPIN1 (also known as PEDF)<sup>23</sup> are unper- turbed in *Dlk1*-mutant NSCs. Moreover, no change was observed in NOTCH activity after addition of recombinant DLK1 to wild-type NSCs, indicating a NOTCH-independent role for DLK1 (Supplementary Fig. 5a–d).

*Dlk1*, which is canonically expressed from the paternally inherited chromosome, belongs to the *Dlk1*–*Dio3* imprinted gene cluster on mouse chromosome 12. Northern blots of RNA from the adult brain confirmed *Dlk1* transcription from the paternal allele (Supplementary Fig. 6a). Moreover, we found very low levels of expression of maternal



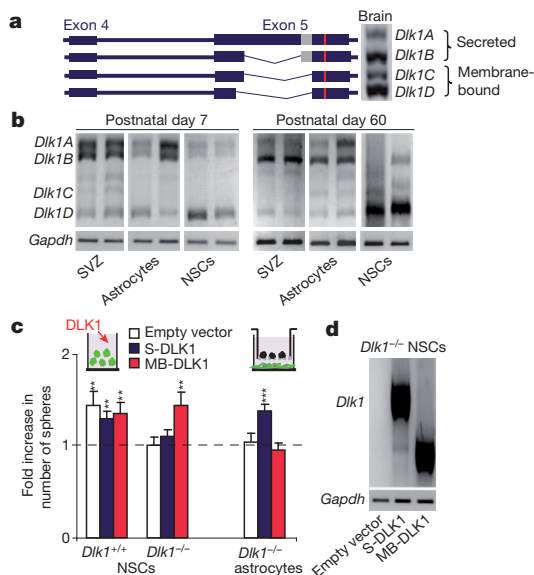
**Figure 2 | DLK1 is secreted by postnatal SVZ niche astrocytes.** **a**, Secondary spheres formed from wild-type and *Dlk1*-mutant primary neurospheres at P7 and P60. **b**, Primary spheres from *Dlk1*<sup>+/+</sup> and *Dlk1*<sup>-/-</sup> SVZ, five days *in vitro* (div) after DLK1 treatment. **c**, Quantification of unipotent (astrocytes, A), bipotent (astrocytes/neurons, A/N) and tripotent (astrocytes/neurons/oligodendrocytes, A/N/O) clones derived from *Dlk1*<sup>+/+</sup> and *Dlk1*<sup>-/-</sup> neurospheres or *Dlk1*<sup>+/+</sup> neurospheres grown in the presence or absence of DLK1 (left panel). Immunocytochemistry for GFAP (red),  $\beta$ III-tubulin (green) and oligodendrocyte marker O4 (yellow) in differentiated neurospheres (right

panel). **d**, Primary neurospheres co-cultured with wild-type or *Dlk1*-mutant niche astrocytes (upper panels). Immunocytochemistry for GFAP (red) and DLK1 (green) in niche astrocyte cultures (lower panels). **e**, Quantification of the P7 and P60 co-culture experiments shown in **d**. **f**, Numbers of primary spheres co-cultured with *Dlk1*<sup>+/+</sup> or *Dlk1*<sup>-/-</sup> astrocytes and treated with DLK1. Dashed lines indicate the numbers of non-co-cultured spheres. \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . All error bars show s.e.m. of triplicate cultures (3–6 samples per genotype). Scale bars: **c**, 50  $\mu$ m; **d**, upper panel, 100  $\mu$ m; lower panel, 30  $\mu$ m.

*Dlk1* in paternal heterozygotes, supporting canonical expression from the paternal allele in septal neurons and ventral striatum (Supplementary Fig. 6b, c). Nonetheless, we observed similar neurogenic phenotypes in *Dlk1*<sup>-/+</sup>, *Dlk1*<sup>+/-</sup> and *Dlk1*<sup>-/-</sup> mutants, and niche astrocytes

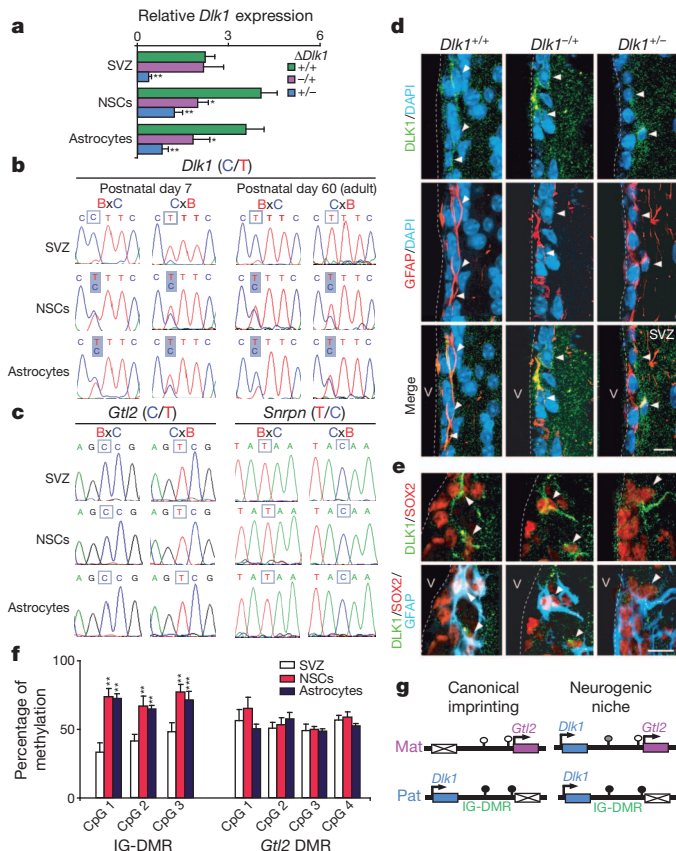
from *Dlk1*<sup>+/-</sup> mice expressed reduced but detectable levels of DLK1 protein (Figs 1d, g and 2d), indicating that *Dlk1* from the maternally inherited allele is required in postnatal neurogenesis. Consistent with the neurogenesis phenotypes, expression was reduced in *Dlk1*<sup>-/+</sup>, *Dlk1*<sup>-/-</sup> and *Dlk1*<sup>+/-</sup> versus *Dlk1*<sup>+/+</sup> NSCs and niche astrocytes, indicating activity from both parental alleles in the neurogenic population (Fig. 4a). We therefore assayed *Dlk1* imprinting in wild-type postnatal mice and adult F<sub>1</sub> hybrid offspring from reciprocal crosses of *Mus musculus domesticus* (C57BL/6J) and *Mus musculus castaneus* (CAST/EiJ) (Supplementary Fig. 6d). SVZ tissue from reciprocal hybrids showed the expected paternal expression of *Dlk1*, whereas neurospheres and niche astrocytes showed biallelic expression of both membrane-bound and secreted isoforms (Fig. 4b and Supplementary Fig. 6e). Importantly, other imprinted genes, such as the adjacent *Gtl2* gene (also known as *Meg3*) and *Snrpn* on chromosome 7C, maintained imprinting in NSCs and niche astrocytes (Fig. 4c). To determine whether biallelic expression was reflected in DLK1 protein levels *in vivo*, immunostaining of DLK1 in combination with GFAP and SOX2 was performed on maternal and paternal heterozygotes. This confirmed that there was expression of DLK1 from both parental chromosomes in the NSCs of the neurogenic zone (Fig. 4d, e). In contrast, postmitotic neurons in non-neurogenic regions showed clear imprinting of *Dlk1* (Supplementary Fig. 6b, c). *Dlk1* was canonically imprinted in NSCs or astrocytes derived from the SVZ of E14.5 embryonic and newborn mice, becoming biallelic by P7 (Supplementary Fig. 6f). These data demonstrate specific and selective absence of *Dlk1* imprinting in the NSC and niche astrocyte populations, commencing at postnatal stages and continuing into adulthood. This indicates that the mechanism conferring postnatal biallelic expression can override the imprint specifically and selectively at *Dlk1*, and that this regulation is required for normal neurogenesis.

*Dlk1*-*Dio3* domain imprinting is controlled by the intergenic differentially methylated region (IG-DMR) located between *Dlk1* and the adjacent non-coding-RNA gene *Gtl2* (refs 2, 24, 25). This germline-derived mark, characterized by hypomethylation on the maternal



**Figure 3 | NSCs require membrane-bound DLK1 to respond to niche-secreted DLK1.** **a**, *Dlk1* transcripts in whole brain. The proteolytic cleavage domain (grey box) is shown in the schematic. **b**, Semi-quantitative PCR of *Dlk1* isoforms in SVZ, NSCs and niche astrocytes at P7 and P60. **c**, Neurospheres generated from *Dlk1*<sup>-/-</sup> or *Dlk1*<sup>+/+</sup> cultures nucleofected with pIRES-GFP, pIRES-GFP-S-DLK1 (secreted isoform) or pIRES-GFP-MB-DLK1 (membrane-bound isoform) after DLK1 treatment, or from wild-type NSCs co-cultured with transduced *Dlk1*<sup>-/-</sup> astrocytes. \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . Error bars show s.e.m. from triplicate cultures ( $n = 6$  animals per genotype). **d**, Semi-quantitative PCR of *Dlk1* isoforms in nucleofected NSCs.





**Figure 4 | NSCs and niche astrocytes selectively lose *Dlk1* imprinting postnatally.** **a**, Quantitative PCR of *Dlk1* expression in SVZ, NSCs and niche astrocytes derived from *Dlk1*<sup>+/+</sup>, *Dlk1*<sup>+/-</sup> and *Dlk1*<sup>-/-</sup> mice. **b**, *Dlk1* allele-specific expression at P7 and P60 in SVZ, NSCs and niche astrocytes derived from reciprocal F<sub>1</sub> hybrid offspring from *Mus musculus domesticus* (C57BL/6, B) and *Mus musculus castaneus* (CAST/Eij, C). **c**, *Gtl2* and *Snrpn* allele-specific expression, as in **b**. **d**, Immunohistochemistry for GFAP (red) and DLK1 (green) in the SVZ of *Dlk1*<sup>+/+</sup>, *Dlk1*<sup>+/-</sup> and *Dlk1*<sup>-/-</sup> mice. **e**, Immunohistochemistry for GFAP (blue), SOX2 (red) and DLK1 (green) in the SVZ of *Dlk1*-mutant mice. Arrowheads in **d** and **e** indicate positive cells. **f**, Methylation at the IG-DMR and *Gtl2*-promoter DMR in SVZ, NSCs and astrocytes. **g**, Schematic representation of the *Dlk1*-*Gtl2* domain in the neurogenic niche. Open and filled circles represent unmethylated and hypermethylated CpGs, respectively. Mat, maternal; Pat, paternal. \*, *P* < 0.05; \*\*, *P* < 0.01. All error bars show s.e.m.; *n* = 10 per group and three bisulphite conversions. Scale bars: **c**, 10  $\mu$ m; **d**, 7  $\mu$ m.

chromosome<sup>25</sup>, is required for the acquisition of differential methylation after fertilization at the secondary *Gtl2*-DMR, which is at the *Gtl2* promoter. Paternal IG-DMR methylation is required for expression of *Dlk1* and repression of *Gtl2* on the paternally inherited chromosome<sup>26</sup>. We measured methylation at these DMRs and found no change at the *Gtl2*-DMR, consistent with retention of its imprinting. In contrast, hypermethylation at the IG-DMR was observed in NSCs and niche astrocytes, indicating that absence of *Dlk1* imprinting is associated with postnatal gain of methylation at the germline DMR (Fig. 4f, g and Supplementary Fig. 7a, b).

Our data support a role for DLK1 in a neurogenic continuum, initiating at the early postnatal period and being maintained over the lifetime of the ageing animal (Supplementary Fig. 8). Compromised neurogenesis seems to be driven by an early increase, followed by a depletion, of the stem-cell pool. In the postnatal SVZ, soluble DLK1 derived from niche astrocytes signals through a membrane-bound form of DLK1 in NSCs to regulate stem-cell number, with a continued requirement for DLK1 in their maintenance with age. Notably, our findings indicate that NSCs and niche astrocytes are distinguished shortly after birth by expression of DLK1 membrane-bound and

secreted isoforms, respectively, and that differential processing of a single gene confers distinct functional properties to these two cell types. Also noteworthy is the epigenetically regulated selective absence of imprinting, resulting in a biallelic dosage of *Dlk1*, which is required for normal neurogenesis. Biallelic expression of imprinted genes in specific cell types may constitute an important regulatory event in their developmental programme. The possibility that loss or gain of genomic imprinting might be used as a dynamic developmental mechanism to control gene dosage has wider implications for our understanding of the function and evolution of imprinting, and indicates that the epigenetic mechanisms that control the process in somatic lineages may be adaptable to the environmental niche in which they are acting.

## METHODS SUMMARY

*Dlk1*-mutant and transgenic mice were maintained as previously described<sup>11,15</sup>. For immunohistochemistry, vibratome sections from brains fixed in 4% paraformaldehyde were used. Neurosphere and primary astrocyte cultures were established as previously described<sup>18</sup>. DLK1 recombinant protein (ENZO Life Sciences) was added to medium at the time of seeding. For co-culture experiments, a transwell insert (Millipore) was used and dissociated NSCs were seeded in the upper compartment. Neurospheres were nucleofected using a mouse NSC nucleofector kit (Amaxa Biosystems), as previously described<sup>23</sup>. Constructs used were pIRES-GFP empty vector (Invitrogen), pIRES-GFP-S-DLK1 or pIRES-GFP-MB-DLK1. Cells were seeded 48 h after nucleofection for neurosphere assessment and immunocytochemistry, as previously described<sup>18</sup>. Antibodies used are listed in Methods. For quantitative PCR using SYBR Green, RNA was isolated with Trizol (Invitrogen) and reverse-transcribed using SuperScript II RT (Invitrogen). Imprinting assays were based on PCR amplification and sequencing of samples derived from reciprocal F1 hybrid offspring of *Mus musculus domesticus* (C57BL/6) and *Mus musculus castaneus* (CAST/Eij). Bisulphite mutagenesis-based cytosine methylation analysis was conducted as described previously<sup>24</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 24 February 2010; accepted 20 May 2011.

- Schmidt, J. V., Matteson, P. G., Jones, B. K., Guan, X. J. & Tilghman, S. M. The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted. *Genes Dev.* **14**, 1997–2002 (2000).
- Takada, S. *et al.* Delta-like and *Gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr. Biol.* **10**, 1135–1138 (2000).
- Riquelme, P. A., Drapeau, E. & Doetsch, F. Brain micro-ecologies: neural stem cell niches in the adult mammalian brain. *Phil. Trans. R. Soc. B* **363**, 123–137 (2008).
- Ma, D. K., Bonaguidi, M. A., Ming, G. L. & Song, H. Adult neural stem cells in the mammalian central nervous system. *Cell Res.* **19**, 672–682 (2009).
- Lledo, P. M., Merkle, F. T. & Alvarez-Buylla, A. Origin and function of olfactory bulb interneuron diversity. *Trends Neurosci.* **31**, 392–400 (2008).
- Song, H., Stevens, C. F. & Gage, F. H. Astroglia induce neurogenesis from adult neural stem cells. *Nature* **417**, 39–44 (2002).
- Doetsch, F. The glial identity of neural stem cells. *Nature Neurosci.* **6**, 1127–1134 (2003).
- Környei, Z. *et al.* Humoral and contact interactions in astroglia/stem cell co-cultures in the course of glia-induced neurogenesis. *Glia* **49**, 430–444 (2005).
- Moon, Y. S. *et al.* Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. *Mol. Cell. Biol.* **22**, 5585–5592 (2002).
- Abdallah, B. M. *et al.* Regulation of human skeletal stem cells differentiation by *Dlk1*/Pref-1. *J. Bone Miner. Res.* **19**, 841–852 (2004).
- Raghunandan, R. *et al.* *Dlk1* influences differentiation and function of  $\beta$  lymphocytes. *Stem Cells Dev.* **17**, 495–508 (2008).
- Davis, E. *et al.* Ectopic expression of DLK1 protein in skeletal muscle of padumal heterozygotes causes the callipyge phenotype. *Curr. Biol.* **14**, 1858–1862 (2004).
- Christophersen, N. S. *et al.* Midbrain expression of Delta-like 1 homologue is regulated by GDNF and is associated with dopaminergic differentiation. *Exp. Neurol.* **204**, 791–801 (2007).
- Bauer, M. *et al.* Delta-like 1 participates in the specification of ventral midbrain progenitor derived dopaminergic neurons. *J. Neurochem.* **104**, 1101–1115 (2008).
- da Rocha, S. T. *et al.* Gene dosage effects of the imprinted delta-like homologue 1 (*Dlk1*/Pref1) in development: implications for the evolution of imprinting. *PLoS Genet.* **5**, e1000392 (2009).
- Li, L., Forman, S. J. & Bhatia, R. Expression of DLK1 in hematopoietic cells results in inhibition of differentiation and proliferation. *Oncogene* **24**, 4472–4476 (2005).
- Tramontin, A. D., García-Verdugo, J. M., Lim, D. A. & Alvarez-Buylla, A. Postnatal development of radial glia and the ventricular zone (VZ): a continuum of the neural stem cell compartment. *Cereb. Cortex* **13**, 580–587 (2003).

18. Ferrón, S. R. *et al.* A combined *ex/in vivo* assay to detect effects of exogenously added factors in neural stem cells. *Nature Protocols* **2**, 849–859 (2007).
19. Kippin, T. E., Martens, D. J. & van der Kooy, D. p21 loss compromises the relative quiescence of forebrain stem cell proliferation leading to exhaustion of their proliferation capacity. *Genes Dev.* **19**, 756–767 (2005).
20. Paik, J. H. *et al.* FoxOs cooperatively regulate diverse pathways governing neural stem cell homeostasis. *Cell Stem Cell* **5**, 540–553 (2009).
21. Bray, S. J., Takada, S., Harrison, E., Shen, S. C. & Ferguson-Smith, A. C. The atypical mammalian ligand Delta-like homologue 1 (Dll1) can regulate Notch signalling in *Drosophila*. *BMC Dev. Biol.* **8**, 11 (2008).
22. Nyfeler, Y. *et al.* Jagged1 signals in the postnatal subventricular zone are required for neural stem cell self-renewal. *EMBO J.* **24**, 3504–3515 (2005).
23. Andreu-Agulló, C., Morante-Redolat, J. M., Delgado, A. C. & Farinas, I. Vascular niche factor PEDF modulates Notch-dependent stemness in the adult subependymal zone. *Nature Neurosci.* **12**, 1514–1523 (2009).
24. Takada, S. *et al.* Epigenetic analysis of the *Dlk1–Gtl2* imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with *Igf2–H19*. *Hum. Mol. Genet.* **11**, 77–86 (2002).
25. Lin, S. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1–Gtl2* imprinted cluster on mouse chromosome 12. *Nature Genet.* **35**, 97–102 (2003).
26. Li, X. *et al.* A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev. Cell* **15**, 547–557 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to J. Herbert and E. B. Keverne for sharing expertise during the course of this work. We thank members of the Ferguson-Smith laboratory for discussions and B. Sun, D. Gray, S. Curran, I. Gutteridge, R. Rancourt and X. d'Anglemont de Tassigny for technical assistance. The work was funded by grants from the Medical Research Council and Wellcome Trust to A.C.F.-S. and by grants from Ministerio de Ciencia e Innovación (SAF2008-01006, CB06/05/0086, RD06/0010/0010) and Generalitat Valenciana (Prometeo) to I.F. S.R.F. is a recipient of a University of Cambridge Herchel-Smith Fellowship.

**Author Contributions** S.R.F. conceived and performed experiments and developed the project, coordinated collaborations and wrote the manuscript. M.C. conducted methylation experiments, helped to design experiments, contributed ideas and interpreted results. E.R. performed expression analysis, contributed to ideas and interpreted results. K.M. helped with imprinting and methylation studies. H.W. and F.G. designed and performed embryo studies. E.H. helped with imprinting and NOTCH-expression studies. J.M.M.-R. performed the NOTCH-pathway experiments. J.L. and S.R.B. provided the *Dlk1*<sup>−/−</sup> mouse. I.F. contributed ideas, discussed results and helped to write the manuscript. A.C.F.-S. conceived and developed the project, designed experiments, interpreted results, coordinated collaborations, contributed ideas and funds and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.C.F.-S. ([afsmith@mole.bio.cam.ac.uk](mailto:afsmith@mole.bio.cam.ac.uk)).

## METHODS

**Animals and *in vivo* manipulations.** The generation and genotyping of *Dlk1*-mutant and transgenic mice have been described previously<sup>11,15</sup>. Except for hybrid analyses, experiments were done with mice on a C57BL/6 genetic background and their corresponding wild-type littermates. Housing of mice and all experiments were carried out in accordance with UK Government Home Office licensing procedures.

**Tissue preparation and immunohistochemistry.** BrdU administration regimens have been previously detailed<sup>18</sup>. Mice at 2–4 months of age were injected intraperitoneally with 50 mg of BrdU per kg of body weight every 2 h for 12 consecutive hours (7 injections in total). Thirty days after the injections, animals were deeply anaesthetized and transcardially perfused with 4% paraformaldehyde (PFA) in 0.1 M phosphate buffer, pH 7.4 (PB), and their brains were vibratome-sectioned at 40  $\mu$ m. The sections were blocked in 10% FBS (v/v, Invitrogen) and 0.1% Triton X-100 (v/v) in PBS (0.9% NaCl in PB) for 1 h, then incubated overnight with primary antibody at 4 °C. For embryonic studies, pregnant females were injected intraperitoneally with a single injection of 100 mg kg<sup>-1</sup> BrdU and embryos (E12.5 and E14.5) were collected 1 h later. Embryonic and postnatal brains were dissected in PBS and fixed for 2 h in 4% PFA at 4 °C. Fixed samples were cryoprotected overnight in 20% sucrose in PBS at 4 °C, mounted in OCT Compound (VWR) and sectioned coronally (20  $\mu$ m) with a cryostat (CM3050S, Leica). For immunohistochemistry, cryostat sections were washed in PBS and blocked at 22 °C for 1 h in PBST (PBS, 0.1% Triton X-100) supplemented with 10% goat serum (Vector Laboratories). The primary antibodies and dilutions used were: mouse antibodies to BrdU (1:250; Dako), MKI67 (1:200; Novocastra),  $\beta$ III-tubulin (1:250; Covance), NESTIN (1:500; Chemicon) or MASH1 (1:100; Beckton Dickinson); goat antibodies to SOX2 (1:100; R&D Systems) or DCX (1:300; Santa Cruz); rabbit antibodies to GFAP (1:500; Dako), S100 $\beta$  (1:100; Dako), TH (1:2,000; Sigma), CALB2 (1:4,000; Swant) or phosphorylated histone H3 (1:500; Upstate); rat antibody to DLK1 (1:100; ENZO Life Sciences). For detection of ependymal cells *in vivo*, sections were labelled with fluorescein isothiocyanate (FITC)-conjugated lectin IB4 (1:1,000; Sigma). For BrdU detection, sections were pre-incubated in 2 M HCl for 30 min at 37 °C and neutralized in 0.1 M sodium borate (pH 8.5). Secondary antibodies were Alexa-488-conjugated donkey anti-mouse, Alexa-488-conjugated anti-rat, Alexa-647-conjugated donkey anti-rabbit (Molecular Probes) and CY3-conjugated donkey anti-goat (Jackson ImmunoResearch Laboratories). These were used at 1:500 for 1 h. Nuclei were counterstained with 1 mg ml<sup>-1</sup> DAPI. For DLK1 staining, *Dlk1*<sup>-/-</sup> SVZ sections or cultures were used to confirm antibody specificity. Immunofluorescence was analysed using a Leica Multispectral confocal microscope (Leica).

For SVZ whole-mounts, we used protocols established previously<sup>27</sup>. Briefly, the lateral walls of the lateral ventricle were dissected out and the resulting whole-mounts were fixed for 1.5 h in 4% paraformaldehyde and washed overnight at 4 °C in PBS. For DCX staining, whole-mounts were washed three times in PBS containing 0.5% Triton X-100 for 15 min each, blocked for 2 h in blocking solution (10% FBS (v/v) and 0.5% Triton X-100 (v/v) in PBS), then incubated for 72 h at 4 °C in goat anti-DCX (1:500; Santa Cruz). The secondary antibody, Alexa-488-conjugated anti-goat (Jackson ImmunoResearch Laboratories), was used at 1:500 dilution. The stained walls were mounted with Fluorsave (Calbiochem) between two coverslips.

**Neural stem-cell and primary astrocyte cultures.** Methods for NSC culture and self-renewal assessment, as well as BrdU immunocytochemistry in neurosphere cultures, have been described previously in detail<sup>18</sup>. Single cells from SVZ dissociates were seeded at very low density (2.5 cells  $\mu$ l<sup>-1</sup>) in neurosphere growth medium with epidermal growth factor (EGF) and fibroblast growth factor (FGF), supplemented with either recombinant DLK1 (mouse) fused to Fc (human) (2–100 ng ml<sup>-1</sup>, produced in HEK-293 cells; ENZO Life Sciences), Jagged 1 (1  $\mu$ g ml<sup>-1</sup>; Calbiochem), or PEDF (50 ng ml<sup>-1</sup>; Bioproducts MD). The cultures were then analysed for formation of primary neurospheres after 5 days *in vitro*. For self-renewal assays, primary spheres formed in any condition were treated with Accutase (0.5 mM; Sigma) for 10 min, mechanically dissociated to a single-cell suspension and re-plated in growth medium containing epidermal growth factor and fibroblast growth factor. Apoptosis and viability in single cells and incipient neurospheres were determined at 24 h and 72 h after plating, as previously described<sup>23</sup>. Multipotency capacity was analysed by seeding individual neurospheres of similar sizes at passage 2–3 in Matrigel-coated 96-well plates for 7 days *in vitro* (with 2% FBS, v/v), before fixation in 4% PFA. No fewer than 50 clones were analysed for each condition and the experiment was conducted as previously described<sup>23</sup>. The distribution of unipotent clones (astrocytes, GFAP<sup>+</sup>), bipotent clones (GFAP<sup>+</sup> astrocytes and  $\beta$ III-tubulin<sup>+</sup> neurons) and tripotent clones (GFAP<sup>+</sup> astrocytes,  $\beta$ III-tubulin<sup>+</sup> neurons and O4<sup>+</sup> oligodendrocytes) was determined. To generate primary astrocyte cultures, we used established methods, as described previously<sup>6,28,29</sup>. Notably, no sub-culturing was made to the astrocyte cultures that were used for co-culture experiments. In brief, isolated adult SVZs

were transferred to Earle's balanced salt solution (EBSS) containing 1.0 mg ml<sup>-1</sup> papain (Worthington DBA), 0.2 mg ml<sup>-1</sup> L-cysteine (Sigma) and 0.2 mg ml<sup>-1</sup> EDTA (Sigma), and incubated in this mixture for 30 min at 37 °C. Tissue was then rinsed in DMEM/F12 medium (1:1 v/v; Invitrogen) and carefully triturated with a fire-polished Pasteur pipette to a single-cell suspension. Isolated cells were collected by centrifugation and resuspended in astrocyte medium (DMEM/F12 medium containing 2 mM L-glutamine, 0.6% glucose, 9.6 g ml<sup>-1</sup> putrescine, 6.3 ng ml<sup>-1</sup> progesterone, 5.2 ng ml<sup>-1</sup> sodium selenite, 0.025 mg ml<sup>-1</sup> insulin and 0.1 mg ml<sup>-1</sup> transferrin, supplemented with 10% FBS). Single-cell suspensions from the SVZ were seeded onto Matrigel-treated wells ( $\times$ 1 Matrigel; Becton Dickinson) in astrocyte medium and cultured at 37 °C with 5% CO<sub>2</sub> until astrocytes reached confluence (approximately 6 days). The medium was changed every 2 days. The flasks were then shaken at 100 rotations min<sup>-1</sup> for 3 h at 22 °C to shake off proliferating cells, oligodendrocyte progenitors and neurons. For co-culture experiments, the astrocyte culture medium was changed to neurosphere growth medium with mitogens, and 2 h afterwards, a transwell insert (0.4- $\mu$ m pore, 12-mm diameter; Millipore) was placed above and dissociated NSCs were seeded in the upper compartment. The astrocyte purity of the cultures was established by immunostaining for GFAP. Primary antibodies and dilutions used *in vitro* were: mouse antibodies to BrdU (1:500; Dako), O4 (1:2, Developmental Studies Hybridoma Bank) or  $\beta$ III-tubulin (1:300; Covance); rabbit antibodies to GFAP (1:700; Dako), NICD (1:200; Abcam) or caspase 3 (1:300; Cell Signaling); rat anti-DLK1 (1:100; ENZO Life Sciences). Secondary antibodies were Alexa-488-conjugated donkey anti-mouse, Alexa-488-conjugated anti-rat, Alexa-647-conjugated donkey anti-rabbit (Molecular Probes) and CY3-conjugated donkey anti-goat (Jackson ImmunoResearch Laboratories). Secondary antibodies were used at 1:500 for 1 h. DAPI (1 mg ml<sup>-1</sup>) was used for counterstaining.

**Cell transduction and luciferase assays.** Constructs encoding the predominant forms of *Dlk1* that are generated by alternate splicing and proteolysis were generated previously<sup>21</sup>. We subcloned them in a pIRESH-GFP vector (Invitrogen). NSCs grown for 2 days (passage 4–6) were nucleofected using a Mouse NSC nucleofector kit (Amaxa Biosystems), as described by the manufacturer, with 2–4  $\mu$ g pIRESH-GFP empty vector, pIRESH-GFP-S-DLK1 (S-DLK1, the secreted form, contains sequences from full-length *Dlk1A* up to the juxtamembrane cleavage domain, amino acid 303), or pIRESH-GFP-MB-DLK1 (MB-DLK1, the membrane-bound form, lacks the coding sequence for the juxtamembrane motif, amino acids 230–303, that is the substrate for cleavage, and so represents *Dlk1C*). Cells were dissociated 2 days after nucleofection and seeded at low density (2.5 cells  $\mu$ l<sup>-1</sup>) in 96-well plates for neurosphere assessment. In reporter assays with a firefly-luciferase-based construct, we electroporated 2–4  $\mu$ g of DNA from the construct 4xwtCBF1-luc<sup>30</sup> and 50 ng of a *Renilla* luciferase construct as an internal control. After 24–36 h, transduced cells (passage 4–6) were dissociated, plated in the presence or absence of DLK1 (50 ng ml<sup>-1</sup>; ENZO Life Sciences) or PEDF (100 ng ml<sup>-1</sup>) as a positive control, then cultured for 24 h before being collected for analysis. Transfection efficiency was around 80% in all cases. Luciferase activity was measured in cell lysates using the Dual Luciferase Assay System (Promega) and promoter activity was defined as the ratio between the firefly and *Renilla* luciferase activities.

**Expression studies.** RNAs were extracted with Trizol (Invitrogen), following the manufacturer's guidelines. For northern blots, mRNA was isolated from 75  $\mu$ g of total RNA using the Dynabeads Oligo (dT)<sub>25</sub> kit (Invitrogen), following the manufacturer's protocol. We used probes described previously for *Dlk1* and *Gapdh*<sup>2</sup>. For quantitative PCR, 1  $\mu$ g of total RNA was DNase-treated with RQ1 RNase-free DNase (Promega), following the manufacturer's guidelines. All cDNA was synthesized using random primers and SuperScript II RT reverse transcriptase (Invitrogen), following standard procedures. Quantitative PCR was used to measure expression levels of *Dlk1* normalized to *Gapdh*, and was performed on a DNA Engine 2 Opticon detection system (Bio-Rad) using SYBR Green I as a double-strand-DNA-specific binding dye. Thermocycling was performed in a final volume of 20  $\mu$ l, containing 2  $\mu$ l of cDNA sample (diluted 1:5), 20 pmol of each primer (*Dlk1-F*, 5'-GAAAGGACTGCCAGCACAAG-3'; *Dlk1-R*, 3'-CACAGAAGTTGCCTGAAGC-5'; *Gapdh-F*, 5'-CCATCACCATCTTCCAGGAG-3'; *Gapdh-R*, 5'-GCATGGACTGTGGTTCATAG-3'), 2 mM MgCl<sub>2</sub>, 0.2 mM dNTP mixture,  $\times$ 1 Taq reaction buffer, 0.5 U HotStart Taq DNA polymerase and 0.5  $\mu$ l of a 1:3,000 dilution of SYBR Green I. Semi-quantitative PCR was performed to determine the different *Dlk1* isoforms, and levels of transcripts were quantified by densitometric analysis of PCR bands in electrophoresis gels stained with ethidium bromide, normalized to *Gapdh* levels obtained at 20 cycles. Primers used were: *Dlk1Splice-F*, 5'-CTGCACACCTGGGTTCTCTG-3'; *DlkSplice-R*, 3'-TCCTCATCACCAGCCTCCTT-5'. *In situ* hybridization was conducted on PFA-fixed 40- $\mu$ m vibratome sections of adult brain, according to procedures previously described<sup>15</sup>.

**Imprinting assays.** All imprinting assays were based on PCR amplification followed by direct sequencing to analyse parental-specific expression of the genes.



PCR reactions were purified with the QiAquick kit (Qiagen) before sequencing. We used reciprocal F1 hybrid offspring of *Mus musculus domesticus* (strain C57BL/6, abbreviated BL6) and *Mus musculus castaneus* (strain CAST/EiJ, abbreviated Cast) subspecies, in which we had identified a single-nucleotide polymorphism between the two subspecies. The sequences of murine *Dlk1* were obtained from GenBank (accession number NM010052). The *Dlk1* polymorphism, located at nucleotide 721 of exon 5, is a 'T' in BL6 mice and a 'C' in Cast mice. The *Dlk1* imprinting assay used the following primers to amplify a fragment between exons 4 and 5: *Dlk1-F*, 5'-ACCCCTGCGCCAACAATGG-3' and *Dlk1-R*, 3'-GGGGTGAAGGGCCTGGGGAGT-5', with the thermal cycler conditions: 94 °C 1 min, 60 °C 1 min, 72 °C 1 min; 30 cycles. The primer *Dlk1seq*, 5'-AGAAGAAGAACCCTCCTGTTGCA-3', was used for direct reverse PCR-fragment sequencing. The sequences of murine *Gtl2* were obtained from GenBank (accession number Y13832). The *Gtl2* polymorphism, located at nucleotide 26 of exon 8, is a 'G' in BL6 mice and an 'A' in Cast mice. The *Gtl2* imprinting assay used the following primers to amplify a fragment between exons 6 and 8: *Gtl2-F*, 5'-CTTGCTGGCCCTGGAGAT-3' and *Gtl2-R*, 3'-AACGTGTTGTGCGTGAAGTC-5', with the thermal cycler conditions: 94 °C 1 min, 59 °C 1 min, 72 °C 1 min; 33 cycles. Primer *Gtl2-R* was used for direct sequencing. The sequences of murine *Snrpn* were obtained from GenBank (accession number NM013670). The *Snrpn* polymorphism, located at nucleotide 1,270 of exon 10, is a 'C' in BL6 mice and a 'T' in Cast mice. The *Snrpn* imprinting assay used the following primers to amplify a fragment between exons 9 and 10: *Snrpn-F*, 5'-CATTATGGCTCCTCCACCTG-3' and *Snrpn-R*, 3'-GTACCTGCAAGCTTTTGTACCC-5', with thermal cycler conditions: 94 °C 1 min, 61 °C 1 min, 72 °C 1 min; 30 cycles. Primer *Snrpn-F* was used for direct sequencing. Genomic DNA sequence traces from BL6 and Cast SVZs were used to identify strain-specific polymorphisms (Supplementary Fig. 6d).

**Combined bisulphite restriction analysis and pyrosequencing.** DNA methylation level was quantified using combined bisulphite restriction analysis and pyrosequencing. The method for bisulphite-based cytosine methylation analysis was adapted from Takada *et al.*<sup>24</sup>. DNA was denatured by adding NaOH to a final concentration of 0.4 M. After 15 min incubation at 50 °C, DNA was treated with 5 M sodium bisulphite and 100 mM hydroxyquinone, and incubated for 18 h at 50 °C. DNA was then desalted with the QiAquick PCR purification kit (Qiagen) according to the manufacturer's instructions, and desulphonated by adding NaOH to a final concentration of 0.3 M and incubating for 15 min at 37 °C. DNA was then neutralized and precipitated by adding 10 µg glycogen, 3 M ammonium acetate and 200 µl absolute ethanol and incubating overnight at -20 °C. To generate the product for each target, two rounds of PCR were done with fully-nested or semi-nested primer pairs. Primers for *Gtl2* DMR (intron 1) were CT-F1 (5'-TGGTTTGGGGGTAGTTTTTTATTGTAG-3') and CT-R1 (5'-AAAAAATACAAATAAATTAATTAACAAATCACAAA-3') for the first PCR. For the second round of PCR, the following primers were used: CT-F2 (5'-ATTTTAAAGATGGTTGATGTTGGGTTT-3') and CT-R1. Primers for

*Gtl2* DMR (promoter) were GPFO (5'-TTTTATTTATAATTAGGGTTTATGTAGGGAAA-3') and GPR3 (5'-TCTTCATTTTACACATCCTTTCAAA-3') for the first PCR, and GPF2 (5'-GTAATTTGTTATAGATTGGGGGGTTT-3') and GPR4 (5'-CTTCAAACAAAAATAACTAACCCTACCAA-3') for the second round of PCR. Primers for the IG-DMR first-round PCR were IGOF (5'-GTGTTAAGGTATATTATGTTAGTGTAGG-3') and IGOR (5'-TACAAACCTTCCTCACTCCAAAAATT-3'), and for the second-round PCR, IGIF (5'-ATATTATGTTAGTGTAGGAAGGATTGTG-3') and IGOR. First and second PCRs were carried out in 25 µl, with 2 U HotStar Taq polymerase (Qiagen), ×1 manufacturer's buffer, 3 mM MgCl<sub>2</sub>, 400 µM dNTPs and 1 µM primers. PCR conditions were: 10 cycles of 95 °C for 40 s, 53 °C for 40 s and 72 °C for 1 min, followed by 30 cycles of 95 °C for 30 s, 53 °C for 30 s and 72 °C for 1 min. The PCR products were digested with appropriate restriction enzymes for combined bisulphite restriction analysis, with each enzyme targeting a representative methylation site. For IG-DMR, NruI, Sau3AI and MluI were used; for *Gtl2*-DMR, HincII and HpyCH4IV and for *Gtl2* intron 1, HhaI and HpyCH4IV. MseI was used in all reactions as a bisulphite-conversion control. The entire process was repeated three times and produced similar results each time. For pyrosequencing analysis, a biotin-labelled primer was used to purify the final PCR product using sepharose beads. The PCR product was bound to Streptavidin Sepharose High Performance (Amersham Biosciences), purified, washed, denatured with 0.2 mol l<sup>-1</sup> NaOH and washed again. Pyrosequencing primer (0.3 µM) was then annealed to the purified single-stranded PCR product and pyrosequencing was performed using the PyroMark Q96MD pyrosequencing system (Qiagen). Primers used were: IG-DMR-F (5'-GTGGTTTGTATG GGTAAGTT-3') and IG-DMR-R (5'-CCCTTCCCTCACTCCAAAAATT-3').

**Statistical analyses.** All statistical tests were performed using the GraphPad Prism Software, version 4.00 for Windows (<http://www.graphpad.com>). The significance of the differences between groups was evaluated in all experiments by analysis of variance, followed by a two-tailed Student's *t*-test. When comparisons were performed with relative values (normalized values and percentages), data were first normalized by using an arcsin transformation. Treatment experiments were analysed by paired *t*-test. Data are presented as the mean ± standard error of the mean (s.e.m.) and the number of experiments performed with independent cultures or animals (*n*) is indicated in the legends.

27. Mirzadeh, Z. *et al.* The subventricular zone en-face: wholemount staining and ependymal flow. *J Vis Exp.* **39**, 1938, doi:10.3791/1938 (2010).
28. Lie, D. C. *et al.* Wnt signalling regulates adult hippocampal neurogenesis. *Nature* **437**, 1370–1375 (2005).
29. García-Marqués, J., De Carlos, J. A., Greer, C. A. & López-Mascaraque, L. Different astroglia permissivity controls the migration of olfactory bulb interneuron precursors. *Glia* **58**, 218–230 (2010).
30. Hsieh, J. J. *et al.* Truncated mammalian Notch1 activates CBF1/RBPJk-repressed genes by a mechanism resembling that of Epstein-Barr virus EBNA2. *Mol. Cell. Biol.* **16**, 952–959 (1996).

# Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors

Pengyu Huang<sup>1</sup>, Zhiying He<sup>1,2</sup>, Shuyi Ji<sup>1</sup>, Huawang Sun<sup>1</sup>, Dao Xiang<sup>2</sup>, Changcheng Liu<sup>1,2</sup>, Yiping Hu<sup>2</sup>, Xin Wang<sup>1,3,4</sup> & Lijian Hui<sup>1</sup>

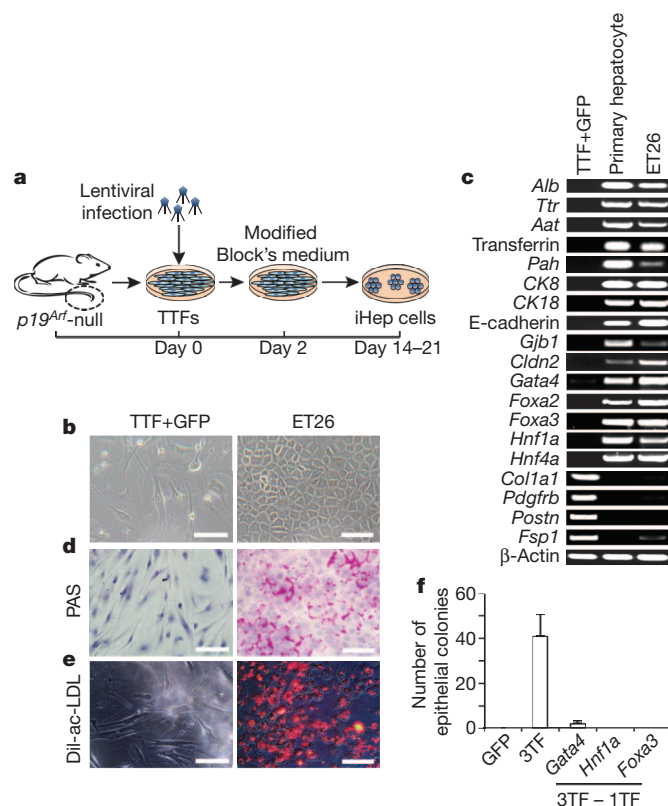
The generation of functional hepatocytes independent of donor liver organs is of great therapeutic interest with regard to regenerative medicine and possible cures for liver disease<sup>1</sup>. Induced hepatic differentiation has been achieved previously using embryonic stem cells or induced pluripotent stem cells<sup>2–8</sup>. Particularly, hepatocytes generated from a patient's own induced pluripotent stem cells could theoretically avoid immunological rejection. However, the induction of hepatocytes from induced pluripotent stem cells is a complicated process that would probably be replaced with the arrival of improved technology. Overexpression of lineage-specific transcription factors directly converts terminally differentiated cells into some other lineages<sup>9–12</sup>, including neurons<sup>13</sup>, cardiomyocytes<sup>14</sup> and blood progenitors<sup>15</sup>; however, it remains unclear whether these lineage-converted cells could repair damaged tissues *in vivo*. Here we demonstrate the direct induction of functional hepatocyte-like (iHep) cells from mouse tail-tip fibroblasts by transduction of Gata4, Hnf1 $\alpha$  and Foxa3, and inactivation of p19<sup>Arf</sup>. iHep cells show typical epithelial morphology, express hepatic genes and acquire hepatocyte functions. Notably, transplanted iHep cells repopulate the livers of fumarylacetoacetate-hydrolase-deficient (*Fah*<sup>−/−</sup>) mice and rescue almost half of recipients from death by restoring liver functions. Our study provides a novel strategy to generate functional hepatocyte-like cells for the purpose of liver engineering and regenerative medicine.

Fourteen mouse transcription factors (14TF, Supplementary Table 1) important for liver development and function<sup>16–19</sup> were transduced into immortalized 3T3 fibroblasts, mouse embryonic fibroblasts (MEFs) and tail-tip fibroblasts (TTFs) via lentiviral infection. The hepatic genes albumin (*Alb*) and *Tdo2* were induced in these cells at day 5 after infection (Supplementary Fig. 1a), indicating that fibroblasts have the potential to be converted to hepatocytes. To ensure that the process is independent of spontaneous immortalization and embryonic progenitors, we focused on TTFs in the following study. Wild-type TTFs showed proliferation arrest and cell death within 7 days after transduction (Supplementary Fig. 1b), thereby inhibiting continuous hepatic conversion. Because p19<sup>Arf</sup> (also called *Cdkn2a*)-null (p19<sup>Arf</sup><sup>−/−</sup>) hepatocytes proliferate *in vitro* without losing genetic stability<sup>20</sup>, we used p19<sup>Arf</sup><sup>−/−</sup> TTFs to overcome the proliferative limitation (Fig. 1a). Remarkably, proliferative cells with epithelial morphology were induced from mesenchymal p19<sup>Arf</sup><sup>−/−</sup> TTFs after transduction of 14TF (Supplementary Fig. 1b). Moreover, these cells expressed *Alb*, *Tdo2* and *Ttr* (Supplementary Fig. 1c).

Eleven epithelial colonies, picked up at day 21 after lentiviral transduction, expressed hepatic genes and the exogenous 14TF at different levels (Supplementary Fig. 2). One epithelial colony, ET26, was further characterized (Fig. 1b). ET26 cells expressed hepatic secretory protein genes, cytokeratin genes, epithelial cell adhesion genes and endogenous hepatic transcription factors (Fig. 1c). By contrast, expression of *Col1a1*, *Pdgfrb*, *Postn* and *Fsp1* (also called *S100a4*), genes typical for fibroblast, was downregulated in ET26 cells (Fig. 1c). Functionally, ET26 cells showed glycogen storage as demonstrated by periodic acid-Schiff

(PAS) staining (Fig. 1d) and uptake of DiI-labelled acetylated low-density lipoprotein (DiI-ac-LDL, Fig. 1e). These results indicated that p19<sup>Arf</sup><sup>−/−</sup> TTFs were converted into cells with significant hepatic gene expression and hepatic functions.

Next, we determined the key factors required for hepatic conversion. On the basis of previous reports<sup>16–19</sup>, we established combinations of six factors (6TF), including Foxa2, Foxa3, Hnf1 $\alpha$ , Hnf4 $\alpha$ , Hnf6 and Gata4, and eight factors (8TF), including 6TF plus Foxa1 and Hnf1. Either 6TF or 8TF converted TTFs to epithelial colonies with hepatic gene expression at comparable levels (Supplementary Fig. 3a, b). Upon withdrawal of Hnf6 from 6TF, we found significantly increased



**Figure 1 | Three transcription factors induce hepatic conversion of tail-tip fibroblasts.** **a**, Experimental design of iHep cell induction. Primary p19<sup>Arf</sup><sup>−/−</sup> TTFs were infected with lentiviruses expressing hepatic transcription factors. Cultures were changed to modified Block's medium 2 days after infection. **b**, Colony ET26 shows epithelial morphology. **c**, Expression of the indicated genes was measured by RT-PCR in ET26 cells, primary hepatocytes and TTFs. **d**, Cytoplasmic accumulation of glycogen was determined by PAS staining (purple cytoplasmic staining). **e**, Intake of DiI-ac-LDL in ET26 cells (red staining). All scale bars: 100  $\mu$ m. **f**, Effects of individual factor withdrawal from 3TF on epithelial colony formation. Data are presented as mean  $\pm$  s.d.

<sup>1</sup>Laboratory of Molecular Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Yueyang Road 320, 200031 Shanghai, China.

<sup>2</sup>Department of Cell Biology, Second Military Medical University, 800 Xiangyin Road, 200433 Shanghai, China. <sup>3</sup>Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, 55455 Minnesota, USA. <sup>4</sup>Stem Cell Institute, University of Minnesota, Minneapolis, 55455 Minnesota, USA.

hepatic gene expression and epithelial colony formation (Supplementary Fig. 3a, b). For the remaining five factors (5TF), removal of *Hnf4 $\alpha$*  further promoted the formation of epithelial colonies (Supplementary Fig. 3c). The remaining four factors were further grouped into two combinations: Gata4, *Hnf1 $\alpha$*  and *Foxa3* (3TF) and Gata4, *Hnf1 $\alpha$*  and *Foxa2* (3TF'). 3TF showed a stronger effect than 3TF' on the induction of hepatic gene expression and epithelial colony formation (Supplementary Fig. 3d and data not shown). Remarkably, 3TF induced endogenous *Foxa2* and *Foxa3* expression (Supplementary Fig. 3d), and removal of any factor from 3TF failed to form epithelial colonies (Fig. 1f). Intriguingly, 3TF triggered *p19<sup>Arf</sup>*<sup>-/-</sup> MEFs to express hepatic genes (Supplementary Fig. 4), indicating the potential to induce hepatic conversion of embryonic fibroblasts. Upon RNA-interference-mediated knockdown of *p19<sup>Arf</sup>*, 3TF also converted wild-type TTFs to epithelial cells with hepatic gene expression (Supplementary Fig. 5).

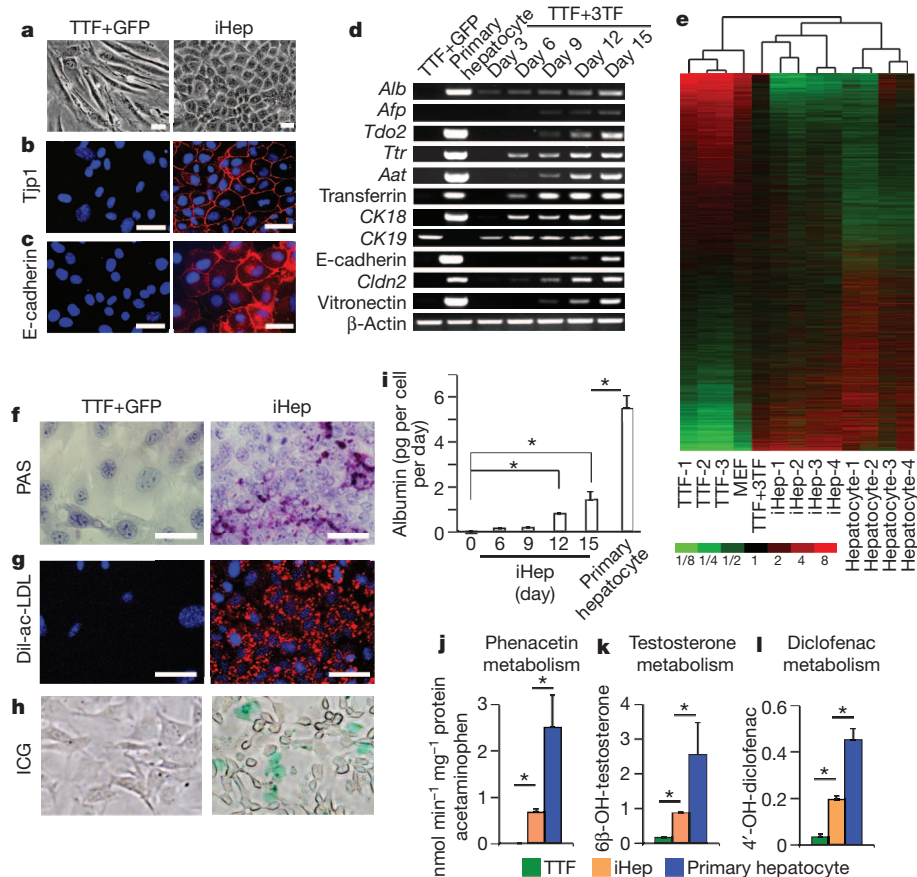
iHep cells induced by the overexpression of Gata4, *Hnf1 $\alpha$*  and *Foxa3* and the inactivation of *p19<sup>Arf</sup>* were characterized for their hepatic features. At day 6, the epithelial cells induced by 3TF were positively stained for tight junction protein 1 (Tjp1) and E-cadherin (Fig. 2a–c). At day 14, 23% of epithelial cells were positive for Alb (Supplementary Fig. 6a), indicating an efficient hepatic conversion. The increased expression of hepatic genes over time, for example, *Alb*, *Ttr*, transferrin (*Ttrf*) and *CK18* (also called *Krt18*), showed a progressively enhanced reprogramming (Fig. 2d and Supplementary Fig. 6b,  $P < 0.05$ ). Interestingly, iHep cells also expressed *Afp* and *CK19* (also called *Krt19*) (Fig. 2d). Protein expression of Alb and *Hnf4 $\alpha$*  was confirmed by immunofluorescent staining in iHep cells (Supplementary Fig. 6c, d). Notably, expression levels of exogenous 3TF were markedly decreased during hepatic conversion, indicating that continuous expression of exogenous 3TF is not required (Supplementary Fig. 6e).

Individual iHep colonies showed similar expression patterns of hepatic genes and fibrotic genes (Supplementary Fig. 6f), indicating

a homogeneous conversion among individual TTFs. Although iHep cells expressed *Afp* and *CK19* (Fig. 2d), other hepatoblast marker genes, such as *Lin28b*, *Igf2* and *Dlk1* (ref. 8), were undetectable during hepatic conversion (Supplementary Fig. 7a). Importantly, cytochrome P450 (CYP) enzymes specific to mature hepatocytes were detectable in iHep cells (Supplementary Fig. 7b), suggesting that hepatic conversion undertakes a process without reversion to progenitors. Moreover, iHep cells neither expressed bile duct marker genes nor formed branching bile duct tubes *in vitro* (Supplementary Fig. 7c, d). The marker genes for pancreatic exocrine and endocrine cells and intestinal cells were also undetectable (Supplementary Fig. 7e, f). Therefore, TTFs are not converted to lineages other than hepatocytes.

We compared the global expression profiles among iHep cells, TTFs, MEFs and hepatocytes cultured for 6 days. Pearson correlation analysis showed that iHep cells were clustered with cultured hepatocytes but separated from TTFs and MEFs (Fig. 2e). Microarray data revealed that numerous hepatic functional genes were upregulated in iHep cells compared to TTFs (Supplementary Figs 8 and 9). When compared with cultured hepatocytes, 877 out of 29,153 annotated genes were found to be upregulated in iHep cells, including *Afp*, *CK19*, *Fabp4* and *S100a9*, whereas 817 genes were downregulated, such as *Cyp4b1*, *Cyp2c40* and *Apob* (fold change  $> 2$ ,  $P < 0.01$ , *t*-test). Notably, iHep cells established substantial hepatic functions. iHep cells accumulated PAS-positive glycogen aggregations and transported DiI-ac-LDL into the cytoplasm (Fig. 2f, g). Indocyanine green uptake was found in 20% of iHep cells (Fig. 2h). Furthermore, iHep cells secreted high amounts of Alb into medium (Fig. 2i,  $P < 0.05$ ). Importantly, iHep cells metabolized phenacetin, testosterone and diclofenac (Fig. 2j–l and Supplementary Fig. 10a–c,  $P < 0.05$ ), whereas metabolic activity for bufuralol was undetected (Supplementary Fig. 10d).

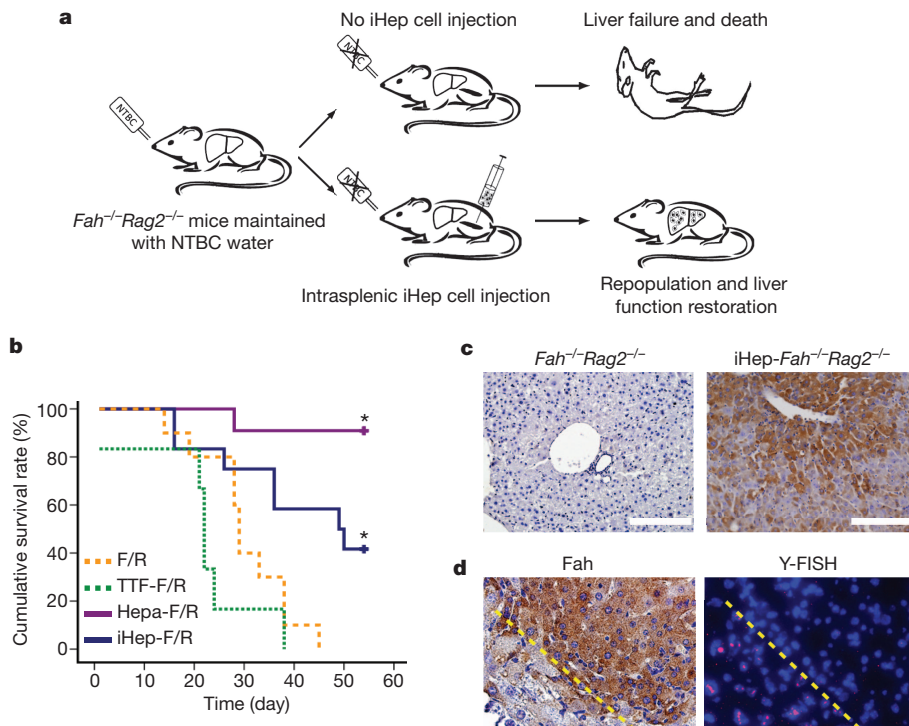
*Fah*<sup>-/-</sup> mice defective in tyrosine metabolism require 2-(2-nitro-4-trifluoro-methylbenzyl)-1,3-cyclohexanedione (NTBC) supply for survival<sup>21–24</sup>. After NTBC withdrawal (NTBC-off), *Fah*<sup>-/-</sup> mice



**Figure 2 | Characterization of iHep cells *in vitro*.**

**a**, 3TF-induced iHep cells show typical epithelial morphology. **b**, **c**, Epithelial conversion of TTFs was confirmed by immunofluorescent staining of Tjp1 and E-cadherin (red membrane staining). Nuclei are stained blue by DAPI. **d**, Expression of indicated genes was analysed by RT-PCR during the induction of iHep cells. **e**, Global gene expression by cDNA microarray assay. Expression profiles were clustered by a Pearson correlation analysis. Expression levels are depicted in colour. TTF+3TF, 3TF-transduced TTFs without enrichment of epithelial cells. Hepatocyte, hepatocytes cultured for 6 days. **f**, Glycogen storage was assayed by PAS staining. **g**, DiI-ac-LDL uptake in iHep cells. **h**, ICG uptake in iHep cells (green staining). **i**, Secretory albumin protein levels were measured by ELISA during hepatic conversion. **j**–**l**, CYP metabolic activities of iHep cells. The metabolic products of phenacetin (converted to acetaminophen by *Cyp1a2*; **j**), testosterone (converted to 6 $\beta$ -OH-testosterone by *Cyp3a* enzymes; **k**) and diclofenac (converted to 4'-OH-diclofenac by *Cyp2c* enzymes; **l**) were determined by liquid chromatography-tandem mass spectrometry according to standard curve. \*,  $P < 0.05$ , *t*-test. All scale bars: 50  $\mu$ m. Data are presented as mean + s.d. in **i**–**l**.

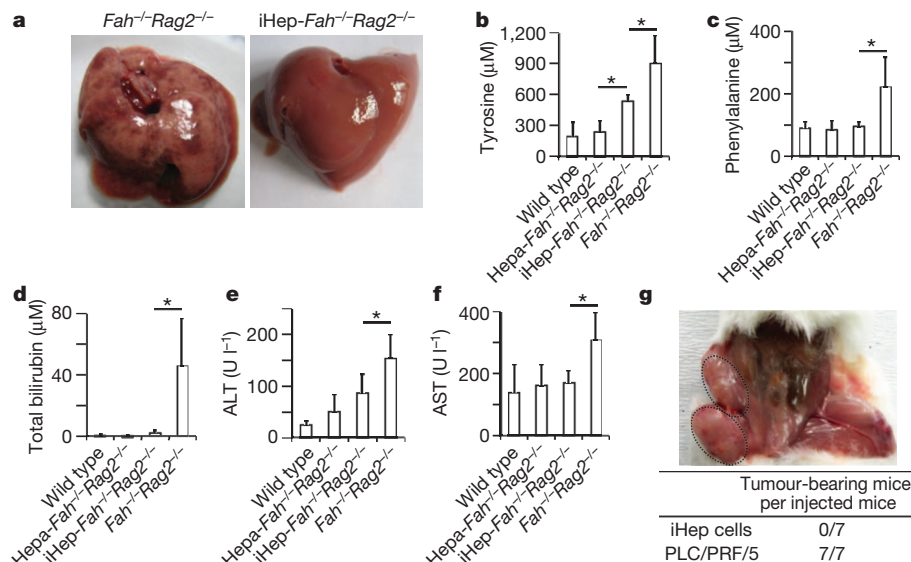




undergo liver failure and death. They can be rescued by transplantation of wild-type primary hepatocytes<sup>21–24</sup>, representing a useful model to characterize *in vivo* repopulation and functions of iHep cells. Immunodeficient *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice were used for transplantation to reduce the likelihood of immunological rejection (Fig. 3a, b and Supplementary Fig. 11a). Ten *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice without transplantation were all dead within 6.5 weeks after NTBC-off and showed continuous loss of body weight (Fig. 3b and Supplementary Fig. 11b). Six *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice transplanted with *p19*<sup>Arf</sup><sup>-/-</sup> TTFs were also dead after NTBC-off (Fig. 3b). In contrast, 5 out of 12 *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice transplanted with iHep cells (iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup>) were alive 8 weeks after NTBC-off and showed increased body weight (Fig. 3b and Supplementary Fig. 11b, *P* < 0.05). *Fah*-positive (*Fah*<sup>+</sup>) iHep cells engrafting into liver sinusoid comprised between 5% to 80% of total hepatocytes in iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> livers (Fig. 3c and Supplementary Fig. 11c). Moreover, *Fah*-wild-type and *p19*<sup>Arf</sup>-null alleles were detected in iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> livers by genomic PCR (Supplementary Fig. 11d). To exclude the possibility of cell fusion between iHep and host

cells, we stained the Y chromosome in male livers transplanted with female iHep cells. Twenty-five *Fah*<sup>+</sup> nodules in four male recipients were characterized and were all negative for Y-chromosome staining, confirming that iHep cells do not fuse with host cells (Fig. 3d and Supplementary Fig. 11e). These results indicate that transplanted iHep cells can repopulate and rescue *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> recipients.

Macroscopically, iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> livers are normal and healthy, whereas livers from NTBC-off *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> control mice were swelled with many necrotic lesions (Fig. 4a). The hexagonal hepatic lobule was destructed due to massive cell death in NTBC-off *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> livers (Supplementary Fig. 12a). In contrast, iHep cell repopulation restored liver architecture without apparent cell death (Supplementary Fig. 12a, b). Remarkably, both repopulated iHep cells and repopulated primary hepatocytes expressed Alb and other hepatic genes at comparable levels in *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> recipient mice (Supplementary Fig. 12c, d). Moreover, serum levels of tyrosine, phenylalanine, ornithine, alanine and glycine were markedly reduced in iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice compared to NTBC-off *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup>



**Figure 4 | iHep cells restore liver functions of *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice.** **a**, Representative pictures of whole livers from *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> and iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice. **b–f**, Serum levels of tyrosine (b), phenylalanine (c), total bilirubin (d), ALT (e) and AST (f) in wild-type (*n* = 6), Hepa-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> (*n* = 5), iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> (*n* = 5, sera collected 8 weeks after iHep transplantation) and control *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice (*n* = 4, sera collected upon losing 20% of body weight). \**P* < 0.05, *t*-test. Data are presented as mean + s.d. **g**, iHep and PLC/PRF/5 cells (human hepatoma cell line) were subcutaneously transplanted into the left and right flanks of NOD/SCID mice, respectively. PLC/PRF/5-generated tumours are indicated by the dotted ovals.

mice (Fig. 4b, c, Supplementary Fig. 12e–g, and Supplementary Table 2,  $P < 0.05$ ). iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice also showed decreased levels of total bilirubin, alanine aminotransferase (ALT) and aspartate aminotransferase (AST) (Fig. 4d–f and Supplementary Table 3,  $P < 0.05$ ). These data demonstrate that iHep cell transplantation substantially improves liver functions of NTBC-off *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice.

Tumours were not found in iHep-*Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> livers 2 months after transplantation. Ki67 staining revealed that iHep cells ceased proliferation 8 weeks after transplantation (Supplementary Fig. 13a). Moreover, iHep cells did not form tumours 8 weeks after subcutaneous xenograft in NOD/SCID mice (Fig. 4g). A total of 20 out of 25 analysed iHep cells displayed 40 chromosomes after 17 passages (Supplementary Fig. 13b), which was comparable with results from wild-type cells (data not shown). These results indicate that iHep cells do not seem to be tumour prone.

To our knowledge, this is the first time that adult fibroblasts have been directly converted to functional iHep cells. Together with previous findings<sup>9–15</sup>, our results prove the general principle that cell lineages can be converted by regulating the transcriptional network. We identified the combination of Gata4, Foxa3 and Hnf1 $\alpha$  as being sufficient to induce hepatic conversion. Gata4 and Foxa3 probably act as ‘pioneer factors’ to trigger a global chromatin modification during hepatic conversion<sup>25,26</sup>. Hnf1 $\alpha$  probably stabilizes the hepatic gene expression, as Hnf1 $\alpha$ , Foxa2 and Hnf4 $\alpha$  occupy each other’s promoters and maintain the hepatic phenotype<sup>16,27</sup>. Moreover, we obtained proliferative iHep cells under the condition of inactivating p19<sup>Arf</sup>. Because p19<sup>Arf</sup> is a key component of the cellular senescence pathway that inhibits induced pluripotent stem cell reprogramming<sup>28</sup>, it would be of interest to characterize whether inactivating other components of this pathway, such as p38 $\alpha$ <sup>29</sup>, could be used to facilitate hepatic conversion.

iHep cells showed an expression profile and hepatic function close to those of mature hepatocytes. Interestingly, some CYP genes were not induced in iHep cells, and *CK19* and *Afp* were upregulated. Moreover, iHep cell transplantation showed that the rescue was partial, suggesting that iHep cells are not identical to hepatocytes. Nevertheless, in contrast with any other cell-type conversion via lineage-specific transcription factors<sup>13–15</sup>, the *in vivo* function of iHep cells has been rigorously proven. More importantly, iHep cells do not seem to be prone to tumour formation. Thus, iHep cells represent an alternative source of hepatocytes for disease modelling, transplantation and tissue engineering. To apply this approach for the purpose of regenerative medicine, future studies will need to address whether human fibroblasts and other cell types could be successfully converted to functional iHep cells.

## METHODS SUMMARY

*p19*<sup>Arf</sup><sup>-/-</sup>, *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> and NOD/SCID mice were maintained according to institutional regulation. *p19*<sup>Arf</sup><sup>-/-</sup> TTFs between passage 7 and 9 were used for induction of iHep cells. *p19*<sup>Arf</sup><sup>-/-</sup> TTFs were seeded on collagen-I-coated dishes and infected with lentiviruses expressing transcription factors. Cells were then cultured in Block’s medium containing 0.1  $\mu$ M dexamethasone, 20  $\mu$ g l<sup>-1</sup> TGF- $\alpha$ , 10  $\mu$ g l<sup>-1</sup> EGF, 4.2 mg l<sup>-1</sup> insulin, 3.8 mg l<sup>-1</sup> human transferrin, and 5  $\mu$ g l<sup>-1</sup> sodium selenite. Fourteen days after infection, we treated cells with 0.01% trypsin and discarded detached fibroblastic cells to enrich the epithelial cells. iHep cells were transplanted into spleens of *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice at the age of 8–12 weeks. We intrasplenically injected  $8.33 \times 10^5$  iHep cells into *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice. NTBC was withdrawn from the drinking water after transplantation. Surviving *Fah*<sup>-/-</sup>*Rag2*<sup>-/-</sup> mice transplanted with iHep cells were killed 8 weeks after the surgery to collect peripheral blood and liver samples. All animal experiments were performed according to institutional regulations. Microarray data have been deposited in the Gene Expression Omnibus database (GSE23635).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 23 August 2010; accepted 15 April 2011.**

**Published online 11 May 2011.**

1. Zern, M. A. Cell transplantation to replace whole liver transplantation. *Gastroenterology* **136**, 767–769 (2009).
2. Gouon-Evans, V. *et al.* BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nature Biotechnol.* **24**, 1402–1411 (2006).

3. Soto-Gutiérrez, A. *et al.* Reversal of mouse hepatic failure using an implanted liver-assist device containing ES cell-derived hepatocytes. *Nature Biotechnol.* **24**, 1412–1419 (2006).
4. Duan, Y. *et al.* Differentiation and enrichment of hepatocyte-like cells from human embryonic stem cells *in vitro* and *in vivo*. *Stem Cells* **25**, 3058–3068 (2007).
5. Basma, H. *et al.* Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. *Gastroenterology* **136**, 990–999 (2009).
6. Si-Tayeb, K. *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51**, 297–305 (2010).
7. Sullivan, G. J. *et al.* Generation of functional human hepatic endoderm from human induced pluripotent stem cells. *Hepatology* **51**, 329–335 (2010).
8. Li, F. *et al.* Hepatoblast-like progenitor cells derived from embryonic stem cells can repopulate livers of mice. *Gastroenterology* **139**, 2158–2169 (2010).
9. Shen, C. N., Slack, J. M. & Tosh, D. Molecular basis of transdifferentiation of pancreas to liver. *Nature Cell Biol.* **2**, 879–887 (2000).
10. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
11. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to  $\beta$ -cells. *Nature* **455**, 627–632 (2008).
12. Feng, R. *et al.* PU.1 and C/EBP $\alpha$ / $\beta$  convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
13. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
14. Ieda, M. *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
15. Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
16. Kyrnizi, I. *et al.* Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev.* **20**, 2293–2305 (2006).
17. Zaret, K. S. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nature Rev. Genet.* **9**, 329–340 (2008).
18. Schrem, H., Klempnauer, J. & Borlak, J. Liver-enriched transcription factors in liver function and development. Part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol. Rev.* **54**, 129–158 (2002).
19. Schrem, H., Klempnauer, J. & Borlak, J. Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation. *Pharmacol. Rev.* **56**, 291–330 (2004).
20. Mikula, M. *et al.* Immortalized p19ARF null hepatocytes restore liver injury and generate hepatic progenitors after transplantation. *Hepatology* **39**, 628–634 (2004).
21. Grompe, M. *et al.* Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. *Genes Dev.* **7** (12A), 2298–2307 (1993).
22. Wang, X. *et al.* The origin and liver repopulating capacity of murine oval cells. *Proc. Natl Acad. Sci. USA* **100** (Suppl. 1), 11881–11888 (2003).
23. Grompe, M. *et al.* Pharmacological correction of neonatal lethal hepatic dysfunction in a murine model of hereditary tyrosinaemia type I. *Nature Genet.* **10**, 453–460 (1995).
24. Overturf, K. *et al.* Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I. *Nature Genet.* **12**, 266–273 (1996).
25. Zaret, K. S. *et al.* Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 119–126 (2008).
26. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
27. Odom, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381 (2004).
28. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460**, 1136–1139 (2009).
29. Hui, L. *et al.* p38 $\alpha$  suppresses normal and cancer cell proliferation by antagonizing the JNK-c-Jun pathway. *Nature Genet.* **39**, 741–749 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank X. Chen, J. Cen, L. Zhang and J. Yuan for technical support, and F. Tang and D. Li for comments. The laboratory of L.H. is funded by the National Science Foundation of China (91019014) and the Chinese Academy of Sciences (KSCX2-YW-R-241, XDA01010402 and the Hundred Talents Program). The laboratory of X.W. is supported by the National Science Foundation of China (30801115), Chinese National 863 Plan Project (2006AA02Z474) and National Key Basic Research and Development Program of China (2007CB947102 and 2009CB941100).

**Author Contributions** L.H. conceived the project. P.H. performed most of the experiments. S.J. analysed the *in vitro* functions of iHep cells. H.S. analysed gene expression of iHep cells. L.H., X.W., P.H. and Z.H. designed the experiments for characterizing *in vivo* functions of iHep cells. P.H., Z.H., D.X., C.L. and Y.H. performed the *in vivo* experiments. L.H. and P.H. analysed the data. L.H., P.H. and X.W. wrote the manuscript.

**Author Information** Microarray data have been deposited in the Gene Expression Omnibus database under accession number GSE23635. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.H. (ljhui@sibs.ac.cn).



## METHODS

**Mice.**  $p19^{Arf-/-}$  mice,  $Fah^{-/-} Rag2^{-/-}$  mice and NOD/SCID mice were maintained in specific pathogen-free husbandry.  $Fah^{-/-} Rag2^{-/-}$  mice were fed with drinking water containing  $7.5 \text{ mg l}^{-1}$  NTBC. The genetic background for  $p19^{Arf-/-}$  and  $Fah^{-/-} Rag2^{-/-}$  mice was C57Bl6/J  $\times$  129Sv.  $Fah^{-/-} Rag2^{-/-}$  mice were used as the recipient to reduce immunological rejection of iHep cells after transplantation.

**Molecular cloning and lentivirus production.** A multi-cloning site (CGGGA TCCCGCGCGCCGACTAGTCGACGCGTCGAGTAACCTACGGACCGGT TT) was inserted into the PmeI restriction site of lentiviral vector pWPI (obtained from Addgene). cDNAs of candidate genes were cloned into the modified pWPI plasmid. For  $p19^{Arf}$  shRNA expression, DNA oligonucleotides encoding  $p19^{Arf}$  shRNA (CCGGGTGAACATGTTGTTGAGGCTAGGATCCTAGCCTCAACA ACATGTTTCACTTTT) were inserted into the AgeI and EcoRI restriction sites of the pLKO.1 plasmid. Constructed pWPI or pLKO.1 plasmids were then introduced to 293FT cells together with packaging plasmid pSPAX2 (Addgene) and envelope plasmid pMD2.G (Addgene). After 48 h incubation, the medium containing lentiviruses was collected and passed through a  $0.45 \mu\text{m}$  filter.

**Fibroblast culture and bile duct induction.** To isolate tail-tip fibroblasts, 5 cm of tail were cut from two-month-old mice. We peeled the dermis and minced tails into 1-cm pieces. Two pieces were placed per 60-mm collagen-I-coated dish in 5 ml DMEM (Sigma-Aldrich) containing 10% FBS (Sigma-Aldrich). After 5 days incubation, fibroblasts that migrated out of the tails were transferred to new collagen-I-coated dishes. We used TTFs between passage 7 and 9 for iHep cell induction. Embryonic fibroblasts were isolated from E13.5 embryos. Head and visceral tissue were dissected and removed. The remaining tissues were minced and incubated with 0.25% trypsin (Gibco) at  $37^\circ\text{C}$  for 15 min. Isolated cells were plated onto a 60-mm collagen-I-coated dish in 5 ml DMEM containing 10% FBS. We used MEFs at passage 3 for lentiviral infection.

For bile duct differentiation,  $1 \times 10^4$  cells were re-suspended in 1 ml DMEM/F12 medium with 1 ml freshly prepared collagen gel solution and poured into a 35-mm dish. After gel solidification, cells were cultured with 1.5 ml DMEM/F12 supplemented with 10% FBS,  $1 \times 10^{-6} \text{ ng ml}^{-1}$  HGF for 3 days.

**Primary hepatocyte isolation and culture.** Adult mice were subjected to standard two-step collagenase perfusion for isolation of primary hepatocytes. Briefly, the liver was pre-perfused through the portal vein with calcium-free buffer (0.5 mM EGTA,  $1 \times$  EBSS without  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ) and then perfused with collagenase ( $0.2 \text{ mg ml}^{-1}$  collagenase type IV (Sigma), 10 mM HEPES,  $1 \times$  EBSS with  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ). Parenchymal cells were purified by Percoll buffer (90% Percoll (Sigma),  $1 \times$  EBSS) at low-speed centrifugation (1,500 r.p.m., 10 min). Viability of isolated hepatocytes was around 90% as determined by Trypan blue. For microarray analysis,  $p19^{Arf-/-}$  primary hepatocytes were cultured in modified Block's medium supplemented with  $0.1 \mu\text{M}$  dexamethasone,  $20 \mu\text{g l}^{-1}$  TGF- $\alpha$ ,  $10 \mu\text{g l}^{-1}$  EGF,  $4.2 \text{ mg l}^{-1}$  insulin,  $3.8 \text{ mg l}^{-1}$  human transferrin and  $5 \mu\text{g l}^{-1}$  sodium selenite in collagen-I-coated dishes for 6 days before harvesting for RNA extraction. For other experiments,  $p19^{Arf-/-}$  primary hepatocytes were immediately lysated in Trizol for total RNA isolation.

**PCR.** For most experiments, total RNA was isolated from cells by Trizol (Invitrogen). For RNA extraction from formalin-fixed-paraffin-embedded (FFPE) tissues, four serial sections mounted on polyethylene terephthalate (PET) membrane frame slides were deparaffinized and air dried. The first section was stained with anti-Fah antibody to identify the repopulated  $Fah^+$  nodules. On the basis of the result of Fah immunostaining in the first section,  $Fah^+$  tissues within the nodules were microdissected from the following three sections by a Leica LMD7000 Laser Microdissection Microscope (Leica Microsystems) with laser intensity of 45 and speed of 5. After microdissection, the remaining sections on the slides were further stained with anti-Fah antibody to confirm that only tissues inside  $Fah^+$  nodules were separated. Microdissected tissues from the same  $Fah^+$  nodule were pooled together for total RNA extraction using RNeasy FFPE Kit (Qiagen).

A total of  $1 \mu\text{g}$  RNA was reverse transcribed into cDNA with M-MLV Reverse Transcriptase (Promega) according to the manufacturer's instructions. For DNA extraction from formalin-fixed-paraffin-embedded tissues, the QIAamp DNA FFPE Tissue Kit (Qiagen) was applied according to the manufacturer's instructions. PCR was performed with HiFi Taq polymerase (TransGen). Quantitative real-time PCR was performed with SYBR Premix Ex Taq (TaKaRa) on an ABI 7500 fast real-time PCR system (Applied Biosystems). Primer sequences will be provided upon request.

**Immunofluorescence.** For immunofluorescence staining, the cells were fixed with 4% paraformaldehyde for 15 min at room temperature, and then incubated with PBS containing 0.2% Triton X-100 (Sigma) for 15 min. Cells were then washed three times with PBS. After being blocked by 3% BSA in PBS for 60 min at room temperature, cells were incubated with primary antibodies at  $4^\circ\text{C}$

overnight, washed three times with PBS, and then incubated with appropriate fluorescence-conjugated secondary antibody for 60 min at room temperature in the dark. Nuclei were stained with DAPI (Sigma). Primary and secondary antibodies were diluted in PBS containing 3% BSA. Antibodies used for immunofluorescence are as follows: mouse anti-Tjp1 (Invitrogen, 1:750), rabbit anti-E-cadherin (Cell Signaling, 1:500), mouse anti-albumin (R&D, 1:200), goat anti-Hnf4 $\alpha$  (Santa Cruz, 1:200). Cy5-conjugated goat anti-mouse IgG (Jackson Laboratories, 1:1,000), Cy3-conjugated goat anti-rabbit IgG (Jackson Laboratories, 1:1,000), Cy3-conjugated donkey anti-goat IgG (Jackson Lab, 1:1,000). For Y-chromosome fluorescent *in situ* hybridization (FISH), liver samples of male  $Fah^{-/-} Rag2^{-/-}$  mice transplanted with female iHep cells were embedded in paraffin and hybridized with mouse Y-chromosome probe (ID Labs Inc., Canada) according to the manufacturer's instruction.

**FACS analyses.** For intracellular staining of albumin,  $10^6$  cells were harvested and fixed with 4% PFA for 30 min, and then permeabilized in staining buffer (PBS with 10% FBS and 0.5% saponin) for 10 min. Cells were then incubated with primary antibody (anti-albumin, R&D) for 30 min in staining buffer, followed with secondary antibody (Cy5-conjugated goat anti-mouse IgG, Jackson Laboratories) incubation for 30 min. Cells were analysed by the Calibur flow cytometer (Becton Dickinson). Data were analysed with Windows Multiple Document Interface for Flow Cytometry (WinMDI, version 2.9).

**PAS stain, DiI-ac-LDL and ICG uptake assays, Alb ELISA and CYP metabolism assay.** Cells were stained by periodic acid-Schiff (PAS, Sigma) and DiI-ac-LDL (Invitrogen) following the manufacturer's instructions. For the indocyanine green (ICG, Sigma) uptake assay, cells were cultured in the medium supplemented with progesterone, pregnenolone-16 $\alpha$ -carbonitrile and 8-bromo cAMP for 2 days. Cells had their medium changed with  $1 \text{ mg ml}^{-1}$  ICG and were incubated at  $37^\circ\text{C}$  for 1 h, followed by washing with PBS three times.

To determine Alb secretion, TTFs transduced with three factors were cultured in the medium without phenol red. Culture supernatant was collected 24 h after medium change. The amount of Alb in the supernatant was determined by the mouse albumin ELISA kit (Bethyl Laboratory) according to the manufacturer's instructions.

For the measurement of CYP enzyme activities, TTFs and iHep cells were cultured in the medium with  $50 \mu\text{M}$  3-methylcholanthrene for 48 h. Cells were dissociated and incubated with substrate in  $200 \mu\text{l}$  incubation medium at different concentrations for 3 h at  $37^\circ\text{C}$ . To stop the reaction,  $800 \mu\text{l}$  cold methanol was added and centrifuged. The supernatants were collected for measurement of indicated productions by LC-MS/MS (Agilent 1200 HPLC and ABI 4000 mass-spectrometer). Freshly isolated hepatocytes were used as a positive control. Total cell protein amount was used to normalize the data. Substrates and metabolic products for standard were commercially purchased: phenacetin, diclofenac, bufuralol, acetaminophen, 4'-OH diclofenac (Sigma), testosterone (Fluka), 6 $\beta$ -OH-testosterone (Cerrilliant) and 1'-OH-bufuralol (Toronto Research Chemicals).

**Microarray analysis.** Total RNA extracted from  $p19^{Arf-/-}$  TTFs,  $p19^{Arf-/-}$  MEFs,  $p19^{Arf-/-}$  hepatocytes cultured for 6 days, 3TF-transduced  $p19^{Arf-/-}$  TTFs without enrichment of epithelial cells, and iHep cells from different experiments was hybridized to whole mouse gene expression microarray (Agilent) under the manufacturer's instruction. Data were normalized by Gene-Spring (Agilent). Microarray hybridization and analysis were carried out by ShanghaiBio Cooperation. Out of 29,153 annotated genes, 11,797 genes for which expression levels were at least twofold different between  $p19^{Arf-/-}$  TTFs and primary  $p19^{Arf-/-}$  hepatocytes were selected for analyses. Hierarchical clustering of samples was performed by Cluster 3.0 software. Average linkage with the uncentred correlation similarity metric was used for the clustering of samples. Original data were uploaded to the Gene Expression Omnibus database (accession number GSE23635).

**In vivo function analysis.**  $Fah^{-/-} Rag2^{-/-}$  mice were maintained with  $7.5 \text{ mg l}^{-1}$  NTBC in the drinking water. According to our previous experience with primary hepatocyte and ES-cell-derived hepatoblast transplantation,  $8.33 \times 10^5$  iHep cells and  $8.33 \times 10^5$   $p19^{Arf-/-}$  TTFs were transplanted into the spleens of  $Fah^{-/-} Rag2^{-/-}$  mice at the age of 8–12 weeks, respectively. NTBC was withdrawn from the drinking water after cell transplantation. Ten  $Fah^{-/-} Rag2^{-/-}$  mice without any transplantation also had NTBC withdrawn as a control. A survival curve was generated by SPSS for windows using Kaplan-Meier method. Eight weeks after transplantation, the blood of surviving iHep-cell-transplanted  $Fah^{-/-} Rag2^{-/-}$  mice was collected from the retro-orbital sinus and centrifuged at 12,000 r.p.m. for 15 min. The serum was frozen at  $-80^\circ\text{C}$  until biochemical analyses. Total bilirubin, albumin, ALT, AST, blood urea nitrogen and creatinine were measured by 7600-020 clinical analyser (Hitachi). Amino acids were quantified by liquid chromatography-mass spectrometry ABI 3200 Q TRAP LC-MS/MS system (Applied Biosystem). After blood collection, mice were killed by cervical dislocation and livers were harvested, fixed and stained with Fah polyclonal antibody or haematoxylin and eosin as previously described. Blood and liver samples of control NTBC-off  $Fah^{-/-} Rag2^{-/-}$  mice were collected after losing 20% body weight.



**Tumour generation assay.** The human hepatoma cell line PLC/PRF/5 was cultured in the same medium as iHep cells. iHep cells were induced and enriched as described above. After 21 days induction, cells were detached by trypsin and suspended in PBS. Seven NOD/SCID mice respectively were injected with  $5 \times 10^6$  iHep cells in the left subcutaneous flank and  $5 \times 10^6$  PLC/PRF/5 cells in the right subcutaneous flank. Tumour numbers were counted 8 weeks after injection.

**Statistics.** All data are presented as mean + s.d. For most statistical evaluation, an unpaired Student's *t*-test was applied for calculating statistical probability in this study. For survival analysis, the Mantel–Cox log-rank test was applied. Statistical calculation was performed using Statistical Program for Social Sciences software (SPSS, IBM). For all statistics, data from at least three independent samples or repeated experiments were used.

# Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors

Sayaka Sekiya<sup>1</sup> & Atsushi Suzuki<sup>1,2</sup>

**The location and timing of cellular differentiation must be stringently controlled for proper organ formation. Normally, hepatocytes differentiate from hepatic progenitor cells to form the liver during development<sup>1,2</sup>. However, previous studies have shown that the hepatic program can also be activated in non-hepatic lineage cells after exposure to particular stimuli or fusion with hepatocytes<sup>3–9</sup>. These unexpected findings suggest that factors critical to hepatocyte differentiation exist and become activated to induce hepatocyte-specific properties in different cell types. Here, by screening the effects of twelve candidate factors, we identify three specific combinations of two transcription factors, comprising *Hnf4a* plus *Foxa1*, *Foxa2* or *Foxa3*, that can convert mouse embryonic and adult fibroblasts into cells that closely resemble hepatocytes *in vitro*. The induced hepatocyte-like (iHep) cells have multiple hepatocyte-specific features and reconstitute damaged hepatic tissues after transplantation. The generation of iHep cells may provide insights into the molecular nature of hepatocyte differentiation and potential therapies for liver diseases.**

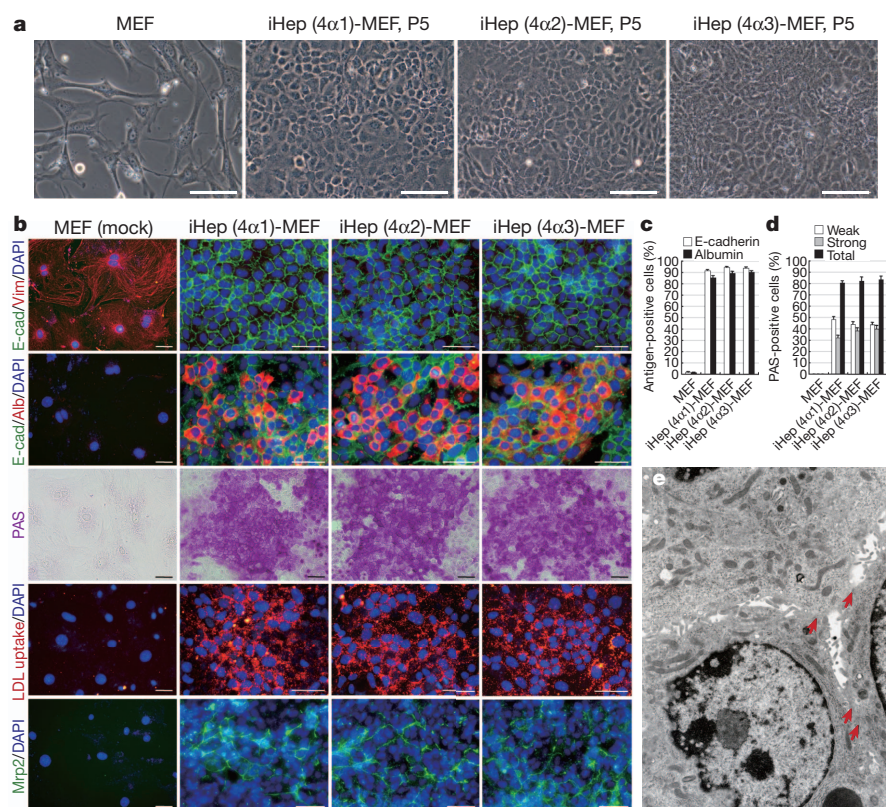
To screen for hepatic fate-inducing factors, we selected 12 candidate genes that are related to hepatocyte differentiation during liver development<sup>1,2</sup>. Retroviruses expressing each gene were prepared, and a mixture of the 12 viruses (referred to as 12MIX) was used to infect mouse embryonic fibroblasts (MEFs) derived from C57BL/6 mice. At 2 weeks after infection with 12MIX, quantitative polymerase chain reaction (qPCR) analyses revealed strong induction of expression of not only the hepatocyte markers albumin and  $\alpha$ -fetoprotein (AFP), but also the epithelial cell marker E-cadherin (also known as *Cdh1*; Supplementary Fig. 1a). To determine the essential factors among the 12 candidate factors, we examined the effects of withdrawing individual factors from the 12MIX pool. Simultaneous reductions of both albumin and AFP expressions were only observed when the viral pool lacked *Hnf4a* (also known as *Hnf4a*), whereas the expression level of E-cadherin was hardly affected by the removal of any of the factors (Supplementary Fig. 1a). Next, we examined the cooperative effects of *Hnf4a* with each of the remaining 11 factors on the expressions of the marker genes. *Hnf4a* elicited its activity in combination with *Foxa1*, *Foxa2* or *Foxa3*, but not in combination with any of the other factors (Supplementary Fig. 1b). Combined expression of *Hnf4a*, *Foxa1*, *Foxa2* and *Foxa3* did not further increase the expression levels of the marker genes (Supplementary Fig. 2). At 2 weeks after infection with individual pools of two factors, comprising *Hnf4a* plus *Foxa1*, *Foxa2* or *Foxa3* (referred to as 4 $\alpha$ 1, 4 $\alpha$ 2 and 4 $\alpha$ 3, respectively), we replated the cells on collagen-coated dishes and continued their culture. Within 3 weeks after replating, morphologically identifiable epithelial-like cells appeared from the fibroblast cultures and proliferated in clusters (Supplementary Fig. 1c). We designated these cells iHep cells and generated three types of iHep cells, namely iHep (4 $\alpha$ 1)-MEFs, iHep (4 $\alpha$ 2)-MEFs and iHep (4 $\alpha$ 3)-MEFs (Fig. 1a). Counting of the numbers of clusters formed by the initial epithelial-like cells showed that 0.3% of MEFs were converted into iHep cells (Supplementary Fig. 3). These iHep cells were successfully maintained in culture with proliferation

(Supplementary Fig. 4a) and showed normal karyotypes (Supplementary Fig. 4b).

Immunofluorescence analyses revealed little or no expressions of the mesenchymal markers vimentin and  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA) in iHep cells (Fig. 1b and Supplementary Fig. 5). In contrast, more than 90% of iHep cells became positive for E-cadherin (Fig. 1b, c and Supplementary Fig. 6a). Albumin was expressed by more than 85% of iHep cells and was coexpressed with E-cadherin (Fig. 1b, c and Supplementary Fig. 6b). Periodic acid–Schiff (PAS) staining revealed glycogen stores in more than 80% of iHep cells (Fig. 1b, d), representing an important function of mature hepatocytes. With regard to hepatocyte properties, iHep cells were competent for low-density lipoprotein (LDL) uptake (Fig. 1b) and expressed the canalicular membrane protein multidrug resistance-associated protein (Mrp) 2 (Fig. 1b), basolateral membrane protein Mrp4 (Supplementary Fig. 7a) and tight junction protein ZO-1 (Supplementary Fig. 7b). Transmission electron microscopy revealed that iHep cells were largely occupied by well-developed ovoid mitochondria, closely attached to adjacent cells by intracellular tight junctional complexes and contained abundant glycogen in their cytoplasm (Fig. 1e). The borders of iHep cells defined luminal spaces that were densely decorated with microvilli, with structures that strongly resembled bile canaliculi between mature hepatocytes (Fig. 1e). Moreover, iHep cells expressed a series of genes encoding liver enzymes, although the expression levels of these genes differed from those in adult mouse hepatocytes (Fig. 2a). In comparisons of the global gene expression profiles of MEFs, iHep cells and adult mouse hepatocytes, iHep cells were clustered closely with hepatocytes but separately from MEFs (Fig. 2b and Supplementary Fig. 8). Indeed, iHep cells mimicked the gene expression patterns of hepatocytes regarding a set of genes involved in fat, cholesterol, glucose and xenobiotic metabolism and genes encoding cytochromes, but differences between iHep cells and hepatocytes were also observed (Fig. 2c and Supplementary Fig. 9). In addition, and similar to hepatocytes, iHep cells secreted albumin (Fig. 2d), produced urea (Supplementary Fig. 10a), yielded glucose (Supplementary Fig. 10b), synthesized triglyceride (Supplementary Fig. 10c), possessed cytochrome P450 activity (Fig. 2e), incorporated and excreted indocyanine green (Fig. 2f) and metabolized drugs (Fig. 2g). Taken together, these findings demonstrate that iHep cells have some of the specific morphological and functional features of hepatocytes. Similar results were obtained for iHep cells generated from BALB/c MEFs (Supplementary Fig. 11).

Next, we sought to characterize iHep cells more precisely. The iHep cells in the small clusters that initially appeared in the MEF cultures already expressed albumin with E-cadherin and contained glycogen stores (Supplementary Fig. 12a), indicating that, in most cases, MEFs were directly converted into cells with hepatocyte properties. Moreover, if iHep cells were bipotent hepatic progenitor cells, cholangiocytes would also be differentiated from iHep cells together with hepatocytes and would easily be observed in culture. However, only a few cells ( $0.24 \pm 0.07\%$ ;  $n = 3$ ) in the iHep cell cultures expressed the cholangiocyte marker cytokeratin (CK) 7 (Supplementary Fig. 12b),

<sup>1</sup>Division of Organogenesis and Regeneration, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan. <sup>2</sup>Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.



**Figure 1 | Generation of iHep cells from MEF cultures and hepatocyte-specific properties in iHep cells.** **a**, Morphologies of primary MEFs and the three types of iHep cells. P, passage number after transduction. **b**, Co-immunofluorescence staining of E-cadherin (E-cad) with vimentin (Vim) or albumin (Alb), PAS staining, LDL uptake assays and immunofluorescence staining of Mrp2 were conducted for mock-infected MEFs and MEF-derived iHep cells. DNA was stained with DAPI. **c**, The percentages of cells immunoreactive for E-cadherin or albumin among MEFs and iHep cells were evaluated by flow cytometry. **d**, The percentages of cells that were weakly or strongly positive for PAS staining among MEFs and iHep cells were calculated after counting ~2,000 cells in individual wells of 12-well plates. **e**, Ultrastructural image of iHep (4α2)-MEFs, original magnification  $\times 11,500$ . The arrows indicate intracellular tight junctional complexes. The data represent means  $\pm$  s.d. ( $n = 3$ ) (**c**, **d**). Scale bars, 100  $\mu\text{m}$  (**a**) and 50  $\mu\text{m}$  (**b**).

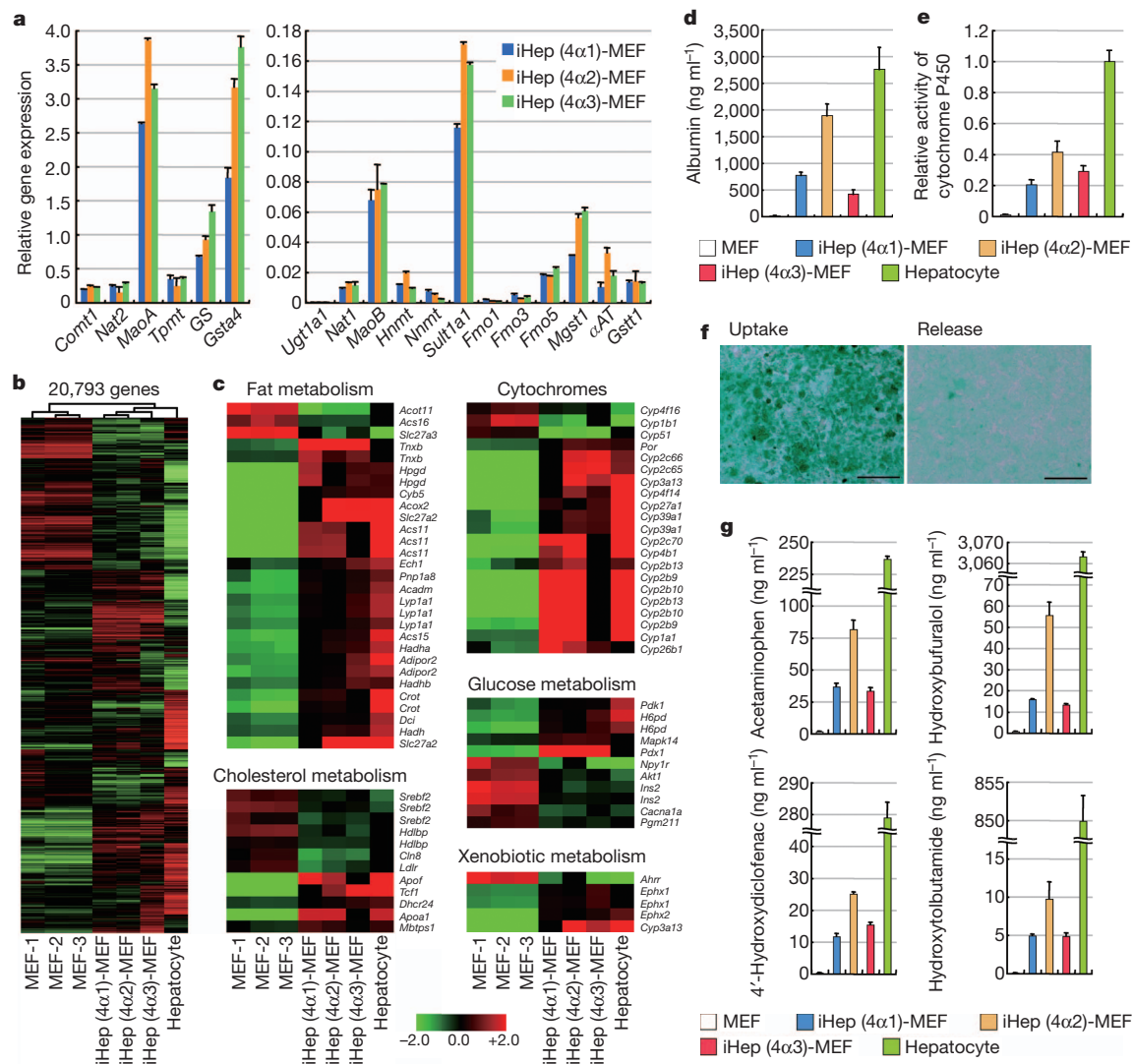
while other cholangiocyte markers were also expressed in the iHep cell cultures (Supplementary Fig. 12c). Single-cell culture analyses of iHep cells showed that  $24.5 \pm 2.6\%$  ( $n = 3$ ) of cells formed clusters of epithelial-like cells that expressed both E-cadherin and albumin and contained glycogen stores, but were not positive for CK7 (Supplementary Fig. 12d). These findings demonstrate that iHep cells are not defined as bipotent hepatic progenitor cells, and that CK7-positive cells may appear with an unexpected bias of transgene expression during the process of iHep cell generation. Moreover, the iHep cell cultures did not contain more primitive endodermal progenitor cells, because there was no expression of markers for pancreatic cells and intestinal cells in the iHep cell cultures (Supplementary Fig. 13). In addition, iHep cells did not require retroviral gene silencing (Supplementary Fig. 14) and became independent of the expression of the exogenous transgenes (Supplementary Fig. 15). One possible reason for the transgene independency of iHep cells was activation of endogenous gene expression. In iHep cells, endogenous *Hnf4α* and *Foxa3* expressions were completely induced, whereas endogenous *Foxa1* and *Foxa2* expressions were also induced but only in some cases (Supplementary Table 1).

Hepatocytes isolated from the adult mouse liver are capable of reconstituting hepatic tissues after transplantation into the livers of fumarylacetoacetate hydrolase (*Fah*)-deficient (*Fah*<sup>−/−</sup>) recipient mice, as a mouse model of hereditary tyrosinaemia type I<sup>10</sup>. Therefore, we examined whether iHep cells can reconstitute hepatic tissues as hepatocytes in the livers of *Fah*<sup>−/−</sup> mice<sup>11</sup>. To this end, we intrasplenically injected the three types of iHep cells, adult mouse hepatocytes and MEFs into *Fah*<sup>−/−</sup> mouse livers. At 1 month after transplantation, iHep cells and hepatocytes, which were both identified as *Fah*-positive hepatocytes, had become engrafted and successfully reconstituted the hepatic tissues in the *Fah*<sup>−/−</sup> recipient mouse livers, whereas no *Fah*-positive cells were observed after transplantation of MEFs (Fig. 3a–e). In the livers of *Fah*<sup>−/−</sup> recipient mice, iHep cells expressed albumin and many of the cells were defined as binucleate mature hepatocytes (Fig. 3f). In addition, iHep cell transplantation ameliorated liver failure, such as increases in bilirubin, alkaline phosphatase (ALP) and alanine transaminase (ALT) and a decrease in serum albumin, similar to the results of hepatocyte

transplantation (Fig. 3g). Survival curves revealed that all of the *Fah*<sup>−/−</sup> mice transplanted with MEFs had died within 27 days after transplantation (Fig. 3h). In contrast, 40% of the *Fah*<sup>−/−</sup> mice transplanted with iHep cells survived for more than 10 weeks, similar to the mice transplanted with hepatocytes (Fig. 3h). At 2 months after transplantation, most of the iHep cells that reconstituted the hepatic tissues in *Fah*<sup>−/−</sup> recipient mouse livers had stopped proliferation, similar to hepatocytes in wild-type mice, as assessed by the expression of the proliferation marker Ki67 (Supplementary Fig. 16a, c). However, these iHep cells were capable of responding to regenerative stimuli after two-thirds partial hepatectomy (PH), and the number of Ki67-positive iHep cells increased to a similar level to Ki67-positive hepatocytes in wild-type mice after PH (Supplementary Fig. 16b, c). Moreover, green fluorescent protein (GFP)-positive cells isolated from the livers of *Fah*<sup>−/−</sup> mice that had been transplanted with GFP-positive iHep cells recapitulated the levels of gene expression in normal hepatocytes (Supplementary Fig. 17). Next, to evaluate the therapeutic potential of iHep cells derived from fibroblasts in mice that are genetically defective in hepatocyte functions, we generated iHep cells from *Fah*<sup>−/−</sup> MEFs (Fig. 3i). The resulting iHep cells, designated iHep (4α3)-*Fah*<sup>−/−</sup> MEFs, expressed E-cadherin and albumin and contained glycogen stores (Supplementary Fig. 18a). We recovered *Fah* expression in iHep (4α3)-*Fah*<sup>−/−</sup> MEFs by infection with a retrovirus coexpressing *Fah* with GFP (Supplementary Fig. 18b–d) and then transplanted these genetically modified iHep cells into *Fah*<sup>−/−</sup> mouse livers. At 1 month after transplantation, donor-derived cells expressing both *Fah* and GFP had become engrafted and successfully reconstituted hepatic tissues in the *Fah*<sup>−/−</sup> mouse livers (Fig. 3j and Supplementary Fig. 18e). These findings demonstrate that iHep cells seem to be morphologically and functionally indistinguishable from hepatocytes after transplantation into the liver, and that genetically modified iHep cells can repair hepatic defects after transplantation.

To examine whether cell fusion occurred, we transplanted iHep cells derived from wild-type female MEFs into *Fah*<sup>−/−</sup> male mouse livers. By combining fluorescence *in situ* hybridization (FISH) with immunofluorescence staining, we did not detect any Y chromosomes in the *Fah*-positive donor-derived cells found in the recipient mouse livers





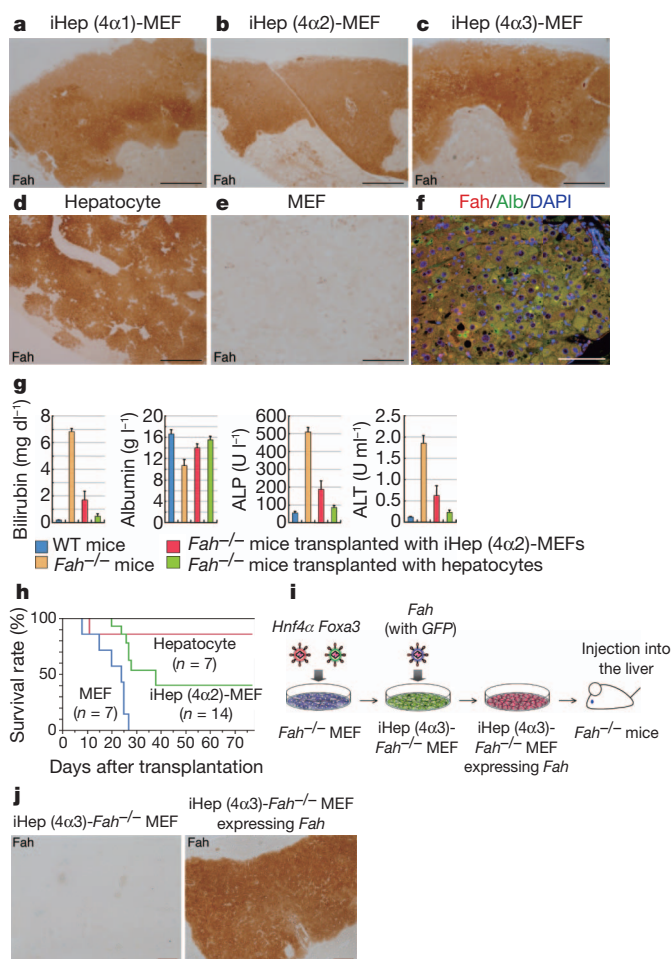
**Figure 2 | Hepatic functions in iHep cells.** **a**, Gene expression analyses by qPCR for iHep cells. **b**, **c**, Global gene expression analyses using microarrays. A hierarchical clustering image reveals differences between MEFs and the other cell samples (**b**). Genes that exhibited significantly different expression levels among genes involved in liver metabolic activity were extracted (values of  $P < 0.05$ ) (**c**). **d**, **e**, The amounts of albumin in the culture media (**d**) and cytochrome P450 activity (**e**) were measured after culture of cells.

(Supplementary Fig. 19a). In addition, we generated iHep cells from fibroblasts of *Alb-Cre* mice expressing Cre recombinase from the albumin genomic locus and iHep cells constitutively expressing Cre recombinase by infection with a virus expressing Cre (Supplementary Fig. 19b). We then transplanted these cells into *Fah*<sup>-/-</sup>;R26R<sup>YFP</sup> mouse livers. When cells expressing Cre recombinase fuse with cells from R26R<sup>YFP</sup> mice<sup>12</sup>, Cre-mediated excision of the floxed termination sequence leads to constitutive yellow fluorescent protein (YFP) expression<sup>13</sup>. After transplantation, the *Fah*<sup>-/-</sup>;R26R<sup>YFP</sup> recipient mouse livers contained *Fah*-positive donor-derived hepatocytes, but no YFP-positive cells (Supplementary Fig. 19c). Therefore, iHep cells have the potential to reconstitute hepatic tissues, without fusion with recipient hepatocytes, after transplantation.

We typically prepared MEFs after removing all the organs in the digestive system, including the liver and intestine. However, it could not be excluded that a small number of cells from these organs contaminated the isolated MEFs and gave rise to iHep cells without lineage conversion of fibroblasts. To examine this possibility, we prepared MEFs and mouse dermal fibroblasts (MDFs) from mouse embryonic limbs and adult mouse skin, respectively, to avoid contamination by

**f**, Indocyanine green uptake and subsequent release by iHep (4α2)-MEFs. Scale bars, 100 μm. **g**, The metabolites of phenacetin, buturalol, diclofenac and tolbutamide in the culture media were quantified after cell culture. All the data shown in **a** and **e** were normalized by the values of hepatocytes or the value of cytochrome P450 activity in hepatocytes, respectively, and the fold differences are shown. The data represent means ± s.d. ( $n = 3$ ) (**a**, **d**, **e**, **g**).

cells from organs of the digestive system. After infection with the viruses expressing *Hnf4α* and *Foxa3*, iHep cells were successfully generated from limb-derived MEFs and MDFs. These iHep cells had hepatocyte-specific properties and reconstituted hepatic tissues in *Fah*<sup>-/-</sup> mouse livers after transplantation (Supplementary Figs 20 and 21). These findings exclude the possibility that iHep cells are derived from cells within organs of the digestive system and confirm that iHep cells are directly induced from fibroblasts. Moreover, *Foxa1* and *Foxa2* were also effective for inducing the conversion of MDFs into iHep cells when these genes were coexpressed with *Hnf4α* (Supplementary Fig. 22). Thus, combined expression of *Hnf4α* with each of the *Foxa* genes is sufficient to convert not only embryonic but also adult mouse fibroblasts into iHep cells. In addition, there is another possibility that iHep cells would be derived from mesenchymal stem cells (MSCs) in the fibroblast cultures. However, we never observed spontaneous conversion of MEFs or MDFs into iHep cells. Moreover, we generated iHep cells from mesenchymal cells isolated from the mouse bone marrow, which should contain enriched MSCs, with similar efficiency to MEFs (Supplementary Fig. 23). These results support the notion that our iHep cells are derived from fibroblasts, and not from MSCs.



**Figure 3 | iHep cells reconstitute hepatic tissues and support hepatic function *in vivo*.** **a–c**, Immunohistochemical staining of Fah at 1 month after transplantation of the three types of iHep cells (**a–c**), adult mouse hepatocytes (**d**) and MEFs (**e**). **f**, Co-immunofluorescence staining of Fah with albumin (Alb) at 1 month after transplantation of iHep (4α2)-MEFs. Yellow indicates merged red and green signals. DNA was stained with DAPI. **g**, The amounts of bilirubin, albumin, ALP and ALT in the plasma of mice. Cell transplantation was conducted at 1 month before the analysis. The data represent means  $\pm$  s.d. ( $n = 3$ ). **h**, Kaplan–Meier survival curves of *Fah*<sup>−/−</sup> mice after cell transplantation. Wilcoxon statistical analyses revealed a significant difference between the curves for iHep (4α2)-MEFs and MEFs ( $P < 0.001$ ), but not between the curves for iHep (4α2)-MEFs and hepatocytes ( $P = 0.294$ ). **i**, Schematic diagram of the experimental procedure. **j**, Immunohistochemical staining of Fah at 1 month after transplantation of iHep (4α3)-*Fah*<sup>−/−</sup> MEFs or iHep (4α3)-*Fah*<sup>−/−</sup> MEFs expressing *Fah*. Scale bars, 500  $\mu$ m (**a–e**) and 100  $\mu$ m (**f, j**).

In the present study, we have shown that combined expressions of only two transcription factors are sufficient to convert fibroblasts into hepatocyte-like cells that can mature to functional hepatocytes *in vivo*. However, it remains unclear why iHep cells, but not parental fibroblasts and adult mouse hepatocytes, are able to proliferate in culture, and whether these factors can generate iHep cells from human somatic cells. Nevertheless, similar to other studies of cell-fate conversion<sup>14–21</sup>, our findings can provide a powerful system not only for studying the molecular nature of cellular identity and plasticity, but also for developing therapeutic strategies for liver diseases.

## METHODS SUMMARY

MEFs and MDFs were grown on gelatine-coated 12-well plates until they reached 20–30% confluency and then incubated in MEF medium (Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 2 mM L-glutamine and penicillin/streptomycin) containing the concentrated viral supernatants and 5  $\mu$ g ml<sup>−1</sup> protamine sulphate for 8 h to overnight. The viral infection was serially repeated five to seven times. At 1 day after the last infection, the medium

was replaced with hepato-medium, comprising a 1:1 mixture of DMEM and F-12, supplemented with 10% FBS, 1  $\mu$ g ml<sup>−1</sup> insulin, 10<sup>−7</sup> M dexamethasone, 10 mM nicotinamide, 2 mM L-glutamine, 50  $\mu$ M  $\beta$ -mercaptoethanol and penicillin/streptomycin. After culture of MDFs or MEFs for 1 or 2 weeks, respectively, the cells were replated on type I collagen-coated six-well plates and grown in hepato-medium containing 20 ng ml<sup>−1</sup> hepatocyte growth factor and 20 ng ml<sup>−1</sup> epidermal growth factor. In cell transplantation studies, donor cells were suspended in 200  $\mu$ l of culture medium and injected intrasplenically into the livers of young to middle-aged *Fah*<sup>−/−</sup> recipient mice (20–25 weeks old). The care of the mice was in accordance with institutional guidelines.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 October 2010; accepted 6 June 2011.

Published online 29 June 2011.

- Zaret, K. S. Regulatory phases of early liver development: paradigms of organogenesis. *Nature Rev. Genet.* **3**, 499–512 (2002).
- Zaret, K. S. & Grompe, M. Generation and regeneration of cells of the liver and pancreas. *Science* **322**, 1490–1494 (2008).
- Scarpelli, D. G. & Rao, M. S. Differentiation of regenerating pancreatic cells into hepatocyte-like cells. *Proc. Natl Acad. Sci. USA* **78**, 2577–2581 (1981).
- Reddy, J. K. et al. Induction and origin of hepatocytes in rat pancreas. *J. Cell Biol.* **98**, 2082–2090 (1984).
- Wang, X., Al-Dhalimy, M., Lagasse, E., Finegold, M. & Grompe, M. Liver repopulation and correction of metabolic liver disease by transplanted adult mouse pancreatic cells. *Am. J. Pathol.* **158**, 571–579 (2001).
- Lee, K. D. et al. *In vitro* hepatic differentiation of human mesenchymal stem cells. *Hepatology* **40**, 1275–1284 (2004).
- Banas, A. et al. Adipose tissue-derived mesenchymal stem cells as a source of human hepatocytes. *Hepatology* **46**, 219–228 (2007).
- Wang, X. et al. Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* **422**, 897–901 (2003).
- Vassilopoulos, G., Wang, P. R. & Russell, D. W. Transplanted bone marrow regenerates liver by cell fusion. *Nature* **422**, 901–904 (2003).
- Overturf, K. et al. Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I. *Nature Genet.* **12**, 266–273 (1996).
- Suzuki, A. et al. Flow cytometric isolation and clonal identification of self-renewing bipotent hepatic progenitor cells in adult mouse liver. *Hepatology* **48**, 1964–1978 (2008).
- Srinivas, S. et al. Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
- Alvarez-Dolado, M. et al. Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. *Nature* **425**, 968–973 (2003).
- Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transcribed cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
- Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663–676 (2004).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to  $\beta$ -cells. *Nature* **455**, 627–632 (2008).
- Feng, R. et al. PU.1 and C/EBP $\alpha$ / $\beta$  convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105**, 6057–6062 (2008).
- Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
- Szabo, E. et al. Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Iwama, H. Miyoshi and R. M. Tanguay for sharing reagents, F. Costantini for providing the *R26<sup>YFP</sup>* mice and E. Gunshima, A. Kaneyuki and H. Kuboyama for excellent technical assistance. This work was supported in part by the Program for Improvement of the Research Environment for Young Researchers from the Special Coordination Funds for Promoting Science and Technology commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, a Grant-in-Aid for Scientific Research from the MEXT of Japan and the Precursory Research for Embryonic Science and Technology Program of the Japan Science and Technology Agency.

**Author Contributions** A.S. designed the study and wrote the paper. S.S. and A.S. performed experiments and analysed and interpreted the data.

**Author Information** Our microarray data have been uploaded to the Gene Expression Omnibus database, accession number GSE29725. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.S. ([suzukicks@bioreg.kyushu-u.ac.jp](mailto:suzukicks@bioreg.kyushu-u.ac.jp)).



## METHODS

**Mice.** C57BL/6 mice (Clea), *Fah*<sup>-/-</sup> mice<sup>11</sup>, *Alb-Cre* mice (Jackson Laboratory), *R26R*<sup>YFP</sup> mice (a gift from F. Costantini) and BALB/c mice (Clea) were used in this study. The experiments were approved by the Kyushu University Animal Experiment Committee, and the care of the animals was in accordance with institutional guidelines.

**Cell culture.** MEFs were prepared from 13.5 days post coitum embryos. The head and visceral tissues were carefully removed from the embryos, and the remaining tissues were minced with a pair of forceps and incubated in a solution containing 2.5 g l<sup>-1</sup> trypsin and 1 mM EDTA (Nacalai Tesque) for 20 min at 37 °C. After the trypsinization, MEF medium (Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 2 mM L-glutamine (Nacalai Tesque) and penicillin/streptomycin (Nacalai Tesque)) supplemented with 25 µg ml<sup>-1</sup> DNase I (Sigma-Aldrich) was added and pipetted to dissociate the tissue fragments. After incubation for 20 min at 37 °C, the triturated cells were collected by centrifugation (400g for 1 min at 4 °C) and resuspended in MEF medium. The cells were plated on 6-cm tissue culture dishes (cells from one embryo per dish) and grown in MEF medium for 3–4 days at 37 °C under 5% CO<sub>2</sub> before freezing. MDFs were prepared from 10-week-old adult mice. The skin behind the ear was obtained using surgical scissors and minced into 5-mm pieces. Small pieces of the tissue were then plated on gelatine-coated 12-well plates and grown in a 1:1 mixture of MF-start medium (Toyobo) and MEF medium for 5–7 days until they reached confluency. Hepatocytes were isolated from 10-week-old adult mouse livers by two-step collagenase digestion<sup>22</sup>. iHep cells and adult mouse hepatocytes were grown in hepato-medium, comprising a 1:1 mixture of DMEM and F-12, supplemented with 10% FBS, 1 µg ml<sup>-1</sup> insulin (Wako), 10<sup>-7</sup> M dexamethasone (Sigma-Aldrich), 10 mM nicotinamide (Sigma-Aldrich), 2 mM L-glutamine, 50 µM β-mercaptoethanol (Nacalai Tesque) and penicillin/streptomycin, containing 20 ng ml<sup>-1</sup> hepatocyte growth factor (Sigma-Aldrich) and 20 ng ml<sup>-1</sup> epidermal growth factor (Sigma-Aldrich). For single-cell culture analyses, cells identified by clone sorting using a FACS Aria (BD Biosciences) were cultured in individual wells of type I collagen-coated 96-well plates (Iwaki), and the clonal colonies formed from each cell were analysed. Mesenchymal cells derived from C57BL/6 mouse bone marrow were obtained from RIKEN BioResource Center and grown in DMEM (low-glucose) containing 10% FBS and penicillin/streptomycin.

**Retrovirus production and transduction of cells.** Mouse *Hex* (also known as *Hhex*), *Gata4*, *Gata6*, *Tbx3*, *Hnf1α* (also known as *Hnf1a*), *Hnf1β* (also known as *Hnf1b*), *Foxa1*, *Foxa2*, *Foxa3*, *Hnf4α*, *Hnf6* (also known as *OneCut1*) and *Fah* cDNAs were obtained by reverse transcription (RT)-PCR using mouse embryonic or adult liver-derived total RNA, and a rat *Cebpa* (also known as *Cebpa*) cDNA was provided by A. Iwama. The cDNAs were subcloned into pGCDNsam and/or pGCDNsam-IRES-GFP, comprising retroviral vectors with a long terminal repeat derived from murine stem cell virus<sup>23</sup>. To produce recombinant retroviruses, plasmid DNA was transfected into 293gp cells (293 cells containing the *gag* and *pol* genes but lacking an *env* gene) along with the VSV-G expression plasmid pCMV-VSV-G (a gift from H. Miyoshi) using linear polyethylenimine (PEI) (Polysciences). At 3 days before transfection, 293gp cells (2 × 10<sup>6</sup>) were plated on poly-L-lysine-coated 10-cm dishes. Meanwhile, 36 µl of 1 mg ml<sup>-1</sup> PEI, 10 µg of retroviral plasmid DNA and 2 µg of pCMV-VSV-G were diluted in 1 ml of DMEM and incubated for 15 min at room temperature. The mixture was then added to the plated 293gp cells in a drop-by-drop manner. After 6 h of incubation at 37 °C under 5% CO<sub>2</sub>, the medium was replaced with fresh MEF medium and the culture was continued. Supernatants from the transfected cells were collected at 24 h after the medium replacement, filtered through 0.2-µm cellulose acetate filters (Sartorius) and concentrated by centrifugation (9,000g for 16 h at 4 °C). The viral pellets were resuspended in Hanks' balanced salt solution (1/140 of the initial supernatant volume). For inducible retroviral gene expression, we used a Retro-X Tet-On Advanced Inducible Expression System (Takara Bio). MEFs and MDFs were grown on gelatine-coated 12-well plates until they reached 20–30% confluency and then incubated in MEF medium containing the concentrated viral supernatants and 5 µg ml<sup>-1</sup> protamine sulphate (Nacalai Tesque) for 8 h to overnight. The viral infection was serially repeated five to seven times. At 1 day after the last infection, the medium was replaced with growth-factor-free hepato-medium. After culture of MDFs or MEFs for 1 or 2 weeks, respectively, the cells were replated on type I collagen-coated six-well plates (Iwaki) and grown in growth factor-containing hepato-medium.

**Gene expression analysis.** qPCR and RT-PCR were conducted as described previously<sup>24,25</sup>. The information regarding the PCR primers and probes was provided in previous reports<sup>25–27</sup>, except for the qPCR primers/probes for E-cadherin (Mm00486909\_g1), catechol-O-methyltransferase 1 (*Comt1*) (Mm00514377\_m1), monoamine oxidase (*Mao*) A (Mm00558004\_m1), *MaoB* (Mm00555412\_m1), thiopurine methyltransferase (*Tpmt*) (Mm01349379\_m1), glutamine synthetase (*GS*, also known as *Glul*) (Mm00725701\_s1), glutathione S-transferase, alpha 4

(*Gsta4*) (Mm00494803\_m1), UDP glucuronosyltransferase 1A1 (*Ugt1a1*) (Mm02603337\_m1), N-acetyltransferase (*Nat*) 1 (Mm00500740\_s1), *Nat2* (Mm00447913\_s1), histamine N-methyltransferase (*Hnmt*) (Mm00475563\_m1), nicotinamide N-methyltransferase (*Nnmt*) (Mm00447994\_m1), sulfotransferase 1A1 (*Sult1a1*) (Mm01132072\_m1), flavin-containing monooxygenase (*Fmo*) 1 (Mm00515795\_m1), *Fmo3* (Mm00514964\_m1), *Fmo5* (Mm00515805\_m1), microsomal glutathione S-transferase 1 (*Mgst1*) (Mm00498294\_m1) and glutathione S-transferase, theta 1 (*Gstt1*) (Mm00492506\_m1) and the RT-PCR primers for exogenous *Hnf4α* (5'-ACAACCTGCTGCAGGAGATGCT-3' and 5'-ACGCACACCGCCTTATTCCAA-3'), exogenous *Foxa2* (5'-ACCTGAAGCCCGAGCACCATT-3' and 5'-ACGCACACCGCCTTATTCCAA-3'), exogenous *Foxa3* (5'-ACTACAGTGCCTGCGAGTCA-3' and 5'-ACGCACACCGCCTTATTCCAA-3'), endogenous *Hnf4α* (5'-CAGGGCTTGGGTGGCATCCT-3' and 5'-CTGCAGGAGCGCGTTGATGGA-3'), endogenous *Foxa1* (5'-CACAGGGTTGGATGTTGTGT-3' and 5'-GTACGCCATGGGACTCATGCA-3'), endogenous *Foxa2* (5'-GGAGCAGCGGCCAGCGAGTTA-3' and 5'-TCTGCTGGATGCCATGGTGA-3') and endogenous *Foxa3* (5'-TGTAGAGAGACCAGCACT-3' and 5'-AGGTCCATGATCCATTGGTA-3'). TaqMan Gene Expression Assay IDs (Applied Biosystems) are shown within parentheses following the names of the genes.

**Karyotype assay.** Karyotypes were determined by quinacrine-Hoechst staining at the International Council for Laboratory Animal Science (ICLAS) monitoring centre in Japan.

**Immunostaining.** Liver tissues were fixed in 20% formalin, dehydrated in ethanol and xylene, embedded in paraffin wax and sectioned. After deparaffinization and rehydration of the sections, antigen retrieval was performed by microwaving in 0.01 M citrate buffer (pH 6.0). For immunohistochemistry, the sections were then incubated with 0.3% hydrogen peroxide in methanol for 20 min at room temperature to quench endogenous peroxidase activity. Cultured cells were washed with phosphate-buffered saline (PBS) and sequentially fixed with 4% paraformaldehyde for 5 min and 25% acetone in methanol for 1 min at room temperature. The fixed cells were washed in PBS containing 0.1% Tween-20 (Nacalai Tesque) and treated with 0.2% Triton X-100 (Nacalai Tesque) for 1 h at room temperature. After washing with PBS/Tween 20 and blocking, the tissue sections and cultured cells were incubated with the following primary antibodies: mouse anti-Fah (1:100; a gift from R. M. Tanguay), rabbit anti-Fah (1:1,000; Abcam), mouse anti-albumin (1:50; R&D Systems), rabbit anti-Ki67 (1:100; Abcam) and rabbit anti-GFP/YFP (1:1,000; MBL) for the tissue sections; and mouse anti-E-cadherin (1:300; BD Biosciences), rabbit anti-albumin (1:3,000; Biogenesis), mouse anti-vimentin (1:1,000; Sigma-Aldrich), mouse anti-α-SMA (1:500; Sigma-Aldrich), mouse anti-Mrp2 (1:50; Enzo Life Sciences), goat anti-Mrp4 (1:250; Abcam), rabbit anti-ZO-1 (1:50; Zymed) and mouse anti-CK7 (1:300; Chemicon) for the cultured cells. After washing, the sections and cells were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (1:500; Dako) specific to the species of the primary antibodies for immunohistochemistry or Alexa 488- and/or Alexa 555-conjugated secondary antibodies (1:200; Molecular Probes) with DAPI for immunofluorescence staining.

**LDL uptake assay.** LDL uptake by cells was assessed by fluorescence microscopy after incubation of the cells with 10 µg ml<sup>-1</sup> acetylated LDL labelled with 1,1'-diiododecyl-3,3',3'-tetramethylindocarbocyanine perchlorate (DiI-Ac-LDL) (Biomedical Technologies) for 4 h at 37 °C and DAPI.

**Measurements of albumin, urea, glucose and triglyceride.** The amounts of mouse albumin and triglyceride in the culture media or cell lysates, respectively, were measured after culture of MEFs and iHep cells for 48 h or adult mouse hepatocytes for 24 h. The amounts of urea in the culture media were measured after the addition of ammonium chloride to cultures of MEFs and iHep cells. The amounts of glucose in the culture media were measured after culture of MEFs, iHep cells and adult mouse hepatocytes for 24 h in a serum-free glucose production medium (pH 7.4) (DMEM without glucose or phenol red (Sigma-Aldrich) supplemented with 2 mM sodium pyruvate (Nacalai Tesque) and 20 mM sodium lactate (Nacalai Tesque)). Mouse albumin, urea, glucose and triglyceride were detected using a Mouse Albumin ELISA Kit (Shibayagi), QuantiChrom Urea Assay Kit (BioAssay Systems), Glucose Assay Kit (Cayman Chemical) and Triglyceride Assay Kit (Cayman Chemical), respectively, according to the corresponding manufacturer's instructions. The absorbance signals were measured with a Multiskan FC microplate reader (Thermo Fisher Scientific).

**Measurements of cytochrome P450 activity.** Cytochrome P450 activity was measured after culture of MEFs, iHep cells and adult mouse hepatocytes for 24 h using a P450-Glo CYP3A4 Assay Kit (Promega), according to the manufacturer's instructions. The luminescent signals were measured with a Luminescencer Octa (ATTO).

**Drug metabolism analyses.** MEFs, iHep cells and adult mouse hepatocytes were grown with 100 µM phenacetin (Nacalai Tesque), 50 µM bufuralol (Sigma-Aldrich),



100  $\mu$ M diclofenac (Sigma-Aldrich) or 50  $\mu$ M tolbutamide (Sigma-Aldrich) for 48 h. The supernatants were collected and mixed with equal volumes of acetonitrile to prevent further enzyme activity. The metabolites, including acetaminophen (Sigma-Aldrich), hydroxybutylfuralol (Sigma-Aldrich), 4'-hydroxydiclofenac (TRC) and hydroxytolbutamide (TRC), were used to make standard curves for the metabolite analyses. The concentrations of the metabolites in the supernatants were measured with a 4000 QTRAP LC/MS/MS system (AB SCIEX).

**Gene expression microarray and data analysis.** Total RNA was prepared from MEFs, iHep cells and adult mouse hepatocytes using an RNeasy Mini Kit (Qiagen). cRNA was amplified and labelled using a Quick Amp Labelling Kit (Agilent Technologies) and hybridized to a 44K 60-mer oligomicroarray (Whole Mouse Genome Microarray Kit; Agilent Technologies) according to the manufacturer's instructions. The hybridized microarray slides were scanned using an Agilent scanner. The relative hybridization intensities and background hybridization values were calculated using Feature Extraction Software version 9.5.1.1 (Agilent Technologies). The raw signal intensities and flags for each probe were calculated from the hybridization intensities and spot information according to the procedures recommended by Agilent Technologies using the Flag criteria in the GeneSpring Software. In addition, the raw signal intensities of two samples were log<sub>2</sub>-transformed and normalized by the quantile algorithm with Bioconductor<sup>28,29</sup>. We selected probes that had the 'P' flag in all seven samples and obtained 20,793 probes as detected genes. In addition, we extracted probes for genes involved in fat, cholesterol, glucose and xenobiotic metabolism and genes encoding cytochromes from probes that had the "P" flag in at least one sample. Heat maps were generated by the MeV software<sup>30</sup>. Normalized intensity values were loaded and adjusted for scaling by the distance from the median of each probe. We used a hierarchical clustering method with Pearson correlation as a distance metric to sort the samples and the genes. To compare up- or down-regulated genes with Venn diagrams from 20,793 genes, ratio (non-log scaled fold-change) and Z-score<sup>31</sup> were calculated. The arithmetic mean of the intensities of three MEF samples (MEFs) was used as a control. The comparisons were c1: MEFs vs hepatocytes, c2: MEFs vs iHep (4 $\alpha$ 1)-MEFs, c3: MEFs vs iHep (4 $\alpha$ 2)-MEFs and c4: MEFs vs iHep (4 $\alpha$ 3)-MEFs. The criteria for the regulated genes: Z-score  $\geq$  2.0 and ratio  $\geq$  2 (upregulated genes), Z-score  $\leq$  -2.0 and ratio  $\leq$  0.5 (downregulated genes). Microarray data analysis was supported by Cell Innovator. Our data have been uploaded to the Gene Expression Omnibus database (accession number GSE29725).

**Cell transplantation and hepatic function test.** iHep cells ( $10^7$ ), adult mouse hepatocytes ( $2 \times 10^6$ ) and MEFs ( $2 \times 10^6$ ) were suspended in 200  $\mu$ l of culture medium and injected intrasplenically into the livers of young to middle-aged *Fah*<sup>-/-</sup> recipient mice (20–25 weeks old). These mice had a shorter lifespan without the provision of 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) (Swedish Orphan International), but were suitable for analysing the function of the transplanted donor cells. We generated all kinds of iHep cells in at least three independent experiments and transplanted them into more than three recipient mice. The iHep cells generated in all experiments were able to become engrafted

and reconstitute the hepatic tissues in the *Fah*<sup>-/-</sup> recipient mouse livers. Since iHep cells were smaller than adult mouse hepatocytes and MEFs, we could increase the number of iHep cells injected into the liver. However,  $2 \times 10^6$  iHep cells were also sufficient to reconstitute the hepatic tissues in the *Fah*<sup>-/-</sup> recipient mouse livers. The *Fah*<sup>-/-</sup> mice were maintained on drinking water containing 7.5 mg l<sup>-1</sup> NTBC, but this treatment was stopped just after transplantation. The amounts of bilirubin, albumin, ALP and ALT in mouse plasma samples were measured using a QuantiChrom Bilirubin Assay Kit (BioAssay Systems), a Mouse Albumin ELISA Kit (Shibayagi), a QuantiChrom Alkaline Phosphatase Assay Kit (BioAssay Systems) and an Alanine Transaminase Activity Assay Kit (Cayman Chemical), respectively, according to the corresponding manufacturer's instructions.

**Western blotting.** Cells were homogenized in lysis buffer comprising 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40 and a protease inhibitor cocktail (Nacalai Tesque). The cell lysates were separated by SDS-PAGE and transferred to Immobilon-P membranes (Millipore). After incubation with primary antibodies against *Fah*, Cre recombinase (Covance) and  $\beta$ -actin (Abcam) at 4 °C overnight with gentle shaking and washing with Tris-buffered saline (50 mM Tris-HCl (pH 7.5) and 150 mM NaCl) containing 0.1% Tween-20, the membranes were incubated with HRP-conjugated secondary antibodies (1:2,000; Dako) specific to the species of the primary antibodies for 2 h at room temperature. Finally, the immune complexes were visualized with ChemiLumi-One (Nacalai Tesque).

22. Seglen, P. O. Hepatocyte suspensions and cultures as tools in experimental carcinogenesis. *J. Toxicol. Environ. Health* **5**, 551–560 (1979).
23. Kaneko, S., Onodera, M., Fujiki, Y., Nagasawa, T. & Nakauchi, H. Simplified retroviral vector GCSap with murine stem cell virus long terminal repeat allows high and continued expression of enhanced green fluorescent protein by human hematopoietic progenitors engrafted in nonobese diabetic/severe combined immunodeficient mice. *Hum. Gene Ther.* **12**, 35–44 (2001).
24. Suzuki, A., Nakauchi, H. & Taniguchi, H. Glucagon-like peptide 1 (1–37) converts intestinal epithelial cells into insulin-producing cells. *Proc. Natl Acad. Sci. USA* **100**, 5034–5039 (2003).
25. Suzuki, A., Iwama, A., Miyashita, H., Nakauchi, H. & Taniguchi, H. Role for growth factors and extracellular matrix in controlling differentiation of prospectively isolated hepatic stem cells. *Development* **130**, 2513–2524 (2003).
26. Suzuki, A. *et al.* Flow cytometric separation and enrichment of hepatic progenitor cells in the developing mouse liver. *Hepatology* **32**, 1230–1239 (2000).
27. Suzuki, A. *et al.* Clonal identification and characterization of self-renewing pluripotent stem cells in the developing liver. *J. Cell Biol.* **156**, 173–184 (2002).
28. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
29. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
30. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).
31. Quackenbush, J. Microarray data normalization and transformation. *Nature Genet.* **32**, 496–501 (2002).

# Protein targeting and degradation are coupled for elimination of mislocalized proteins

Tara Hessa<sup>1</sup>, Ajay Sharma<sup>1</sup>, Malaiyalam Mariappan<sup>1</sup>, Heather D. Eshleman<sup>1†</sup>, Erik Gutierrez<sup>1,2</sup> & Ramanujan S. Hegde<sup>1†</sup>

A substantial proportion of the genome encodes membrane proteins that are delivered to the endoplasmic reticulum by dedicated targeting pathways<sup>1</sup>. Membrane proteins that fail targeting must be rapidly degraded to avoid aggregation and disruption of cytosolic protein homeostasis<sup>2,3</sup>. The mechanisms of mislocalized protein (MLP) degradation are unknown. Here we reconstitute MLP degradation *in vitro* to identify factors involved in this pathway. We find that nascent membrane proteins tethered to ribosomes are not substrates for ubiquitination unless they are released into the cytosol. Their inappropriate release results in capture by the Bag6 complex, a recently identified ribosome-associating chaperone<sup>4</sup>. Bag6-complex-mediated capture depends on the presence of unprocessed or non-inserted hydrophobic domains that distinguish MLPs from potential cytosolic proteins. A subset of these Bag6 complex 'clients' are transferred to TRC40 for insertion into the membrane, whereas the remainder are rapidly ubiquitinated. Depletion of the Bag6 complex selectively impairs the efficient ubiquitination of MLPs. Thus, by its presence on ribosomes that are synthesizing nascent membrane proteins, the Bag6 complex links targeting and ubiquitination pathways. We propose that such coupling allows the fast tracking of MLPs for degradation without futile engagement of the cytosolic folding machinery.

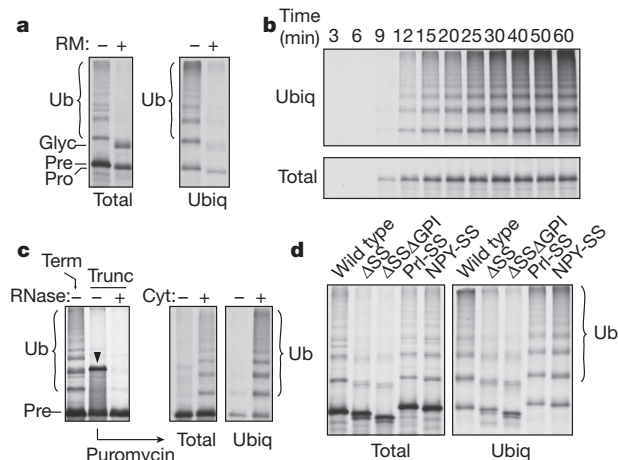
Protein targeting and translocation to the endoplasmic reticulum (ER) are not perfectly efficient<sup>5,6</sup>, thereby necessitating pathways for the degradation of MLPs that have been inappropriately released into the cytosol. For example, mammalian prion protein (PrP), a widely expressed glycosyl phosphatidylinositol (GPI)-anchored cell surface glycoprotein, displays ~5–15% translocation failure *in vitro* and *in vivo*<sup>2,3,9–10</sup>. This non-translocated population of PrP is degraded efficiently by a proteasome-dependent pathway, limiting the cytosolic PrP levels at steady state<sup>2,3,9,10</sup>. Prompt degradation is essential because mislocalized PrP can aggregate, make inappropriate interactions, and cause cell death and neurodegeneration<sup>2,11–14</sup>. The pathways for efficient disposal of MLPs, however, are not known.

To study this problem, we reconstituted the ubiquitination of mislocalized PrP *in vitro*. Radiolabelled PrP synthesized in rabbit reticulocyte lysate (RRL) supplemented with ER-derived rough microsomes was predominantly translocated into the ER, processed and glycosylated (Fig. 1a). However, various conditions that reduced the extent of translocation—such as omission of rough microsomes, inactivation of signal recognition particle (SRP)-dependent targeting or blocking of translocation through the translocon—all resulted in increased PrP ubiquitination in a lysine-dependent manner (Fig. 1a and Supplementary Figs 1–3). Other mislocalized secretory and membrane proteins were also similarly ubiquitinated in the cytosol (Supplementary Fig. 4). The ubiquitination of mislocalized PrP closely parallels PrP synthesis (Fig. 1b), suggesting that ubiquitination is rapid. Yet, ubiquitination occurred strictly post-translationally, because full-length PrP that was tethered as a nascent peptidyl-transfer RNA to the ribosome was not ubiquitinated until it had been released into the cytosol through the

action of puromycin (Fig. 1c and Supplementary Fig. 5). An unrelated membrane protein behaved similarly (Supplementary Fig. 6).

Efficient ubiquitination of PrP was strongly dependent on unprocessed hydrophobic signals at the amino and carboxy termini (Fig. 1d). Conversely, green fluorescent protein (GFP) became a substrate for ubiquitination when hydrophobic targeting signals were added (Supplementary Fig. 4). Ubiquitination was therefore not solely a consequence of protein misfolding, because PrP lacking both the N-terminal targeting signal (denoted  $\Delta$ SS) and the C-terminal GPI-anchoring signal ( $\Delta$ GPI) was misfolded owing to its lack of glycosylation and disulphide bond formation, but was poorly ubiquitinated. This finding suggested the existence of a specialized pathway for hydrophobic-domain-containing MLPs that works more rapidly than traditional quality control pathways, which engage only after repeated failures at folding<sup>15,16</sup>.

To identify factors involved in the MLP degradation pathway, we combined biochemical fractionation and functional reconstitution assays. We produced a translation-competent fractionated RRL (Fr-RRL) (Supplementary Fig. 7) that selectively decreased the ubiquitination of non-translocated PrP (Fig. 2a) and other MLPs (Supplementary



**Figure 1 | Non-translocated PrP is rapidly ubiquitinated.** **a**, The translation of radiolabelled PrP in RRL, with or without rough microsomes (RMs), was analysed directly (left) or after isolation of ubiquitinated (ubiq) products (right) by using SDS-PAGE and autoradiography. Glycosylated (glyc), precursor (pre), processed (pro) and ubiquitinated (Ub) bands are indicated. **b**, Time course of PrP synthesis (bottom) and PrP ubiquitination (top) *in vitro*. **c**, PrP containing a termination codon (term) or lacking this codon (trunc) was translated *in vitro*. Truncated PrP was released using puromycin, in the absence or presence of cytosol (cyt), and total protein and ubiquitination were analysed. The arrowhead indicates tRNA-containing PrP, which can be digested by RNase. **d**, Wild-type PrP or constructs lacking the signal sequence ( $\Delta$ SS) or both the signal sequence and GPI anchor ( $\Delta$ SSAGPI) were analysed directly or after isolation of ubiquitinated products. PrL-SS and NYP-SS contain signal sequence from prolactin and neuropeptide Y, respectively.

<sup>1</sup>Cell Biology and Metabolism Program, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>†</sup>Present addresses: Department of Physiology, University of California, San Francisco, California 94158, USA (H.D.E.); MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK (R.S.H.).

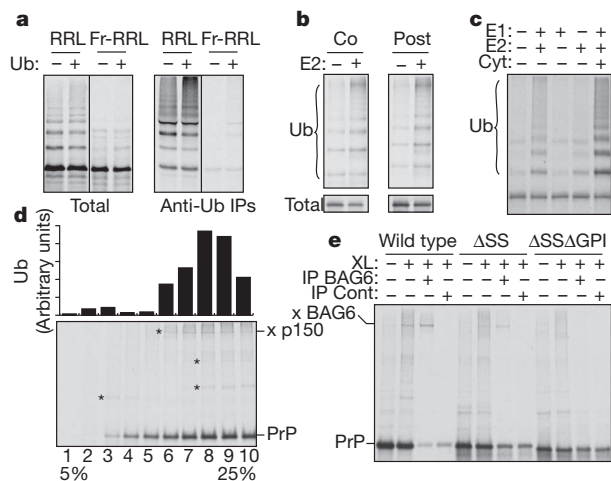
Fig. 8) but not ubiquitination in general (Supplementary Fig. 7). The missing factor in Fr-RRL (other than ubiquitin, which we included in all assays) proved to be the E2 ubiquitin-conjugating enzyme UBC5 (also known as UBE2D1) (Fig. 2b and Supplementary Figs 8 and 9). Because UBC5 restored ubiquitination equally well when added during or after PrP translation (Fig. 2b), we surmised that at least a certain population of PrP remains in a ubiquitination-competent state. Indeed, PrP and other MLPs that were affinity purified from Fr-RRL under native conditions could be ubiquitinated simply by adding purified E1, UBC5, ubiquitin and ATP (Fig. 2c and Supplementary Fig. 10).

To identify factors that maintain the ubiquitination competence of MLPs, the Fr-RRL translation products were separated by size in a sucrose gradient, and each fraction was subjected to parallel ubiquitination and chemical crosslinking analyses (Fig. 2d and Supplementary Fig. 11). The fractions retaining maximum ubiquitination competence for two different substrates correlated well with a ~150-kDa crosslinking partner (Fig. 2d and Supplementary Fig. 11). This interaction was direct (Supplementary Fig. 12) and was strongly dependent on the presence of unprocessed N- and C-terminal signals in PrP (Fig. 2e and Supplementary Fig. 13), correlating with the requirements for ubiquitination (Fig. 1d). On the basis of molecular weight, dependence on hydrophobic domains for interaction and migration position in the sucrose gradient, we surmised that the ~150-kDa crosslinked protein might be BAG6 (also called BAT3 and Scythe), a hypothesis that was subsequently verified by immunoprecipitation experiments (Fig. 2e and Supplementary Figs 13 and 14). BAG6 was recently identified as part of a three-protein ribosome-interacting chaperone complex (composed of BAG6, TRC35 and UBL4A)<sup>4</sup> that is involved in tail-anchored membrane-protein insertion into the ER<sup>4,17</sup>. A combination of crosslinking, affinity purification and immunoblotting studies verified that all three subunits of this complex are associated with MLPs

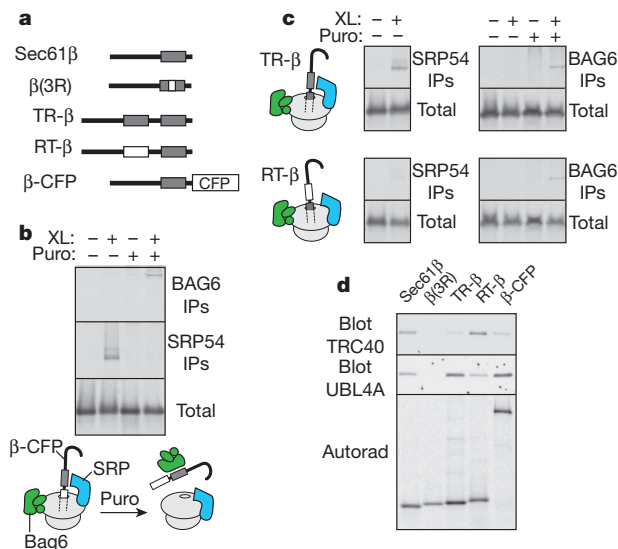
(Supplementary Figs 14 and 15, and data not shown). Thus, the Bag6 complex binds to multiple MLPs through their hydrophobic domains and has a broader specificity than only binding tail-anchored proteins.

To determine when the Bag6 complex first captures MLPs, we analysed ribosome-nascent chains (RNCs) of membrane proteins. When a transmembrane domain (TMD) emerged from the ribosomal 'tunnel', a direct interaction with SRP54 (the signal-sequence-binding subunit of the SRP) was detected by crosslinking experiments (Fig. 3a–c). By contrast, the Bag6 complex, even though it has been found to reside on such RNCs and is abundant in the cytosol<sup>4</sup>, did not make direct contact with the substrate (Fig. 3b, c). When the TMD was still inside the ribosomal tunnel, the RNC was not crosslinked to either BAG6 or SRP54 (Fig. 3c), even though both complexes can be recruited to such ribosomes<sup>4,18</sup>. After puromycin release of each of these RNCs (with the TMD inside versus outside the ribosomal tunnel), BAG6 crosslinking was observed (Fig. 3b, c). Thus, the Bag6 complex captures substrates concomitant with or after the release of nascent chains from the ribosome; these same hydrophobic domains are bound by the SRP as long as the TMD is exposed as an RNC<sup>19</sup>.

Earlier analysis of tail-anchored and non-tail-anchored membrane proteins had shown that only tail-anchored membrane proteins are efficiently loaded onto TRC40 (also known as ASNA1), the targeting factor for tail-anchored protein insertion into the ER<sup>20</sup>. Indeed, modifying a tail-anchored protein either by placing cyan fluorescent protein (CFP) polypeptide sequences after the TMD (a construct denoted  $\beta$ -CFP) (Fig. 3a) or by adding an extra TMD (denoted TR- $\beta$ ) reduced the interactions with TRC40 and simultaneously increased the interactions with the Bag6 complex (Fig. 3d). Similarly, comparison of the crosslinking partners of PrP and those of the tail-anchored protein Sec61 $\beta$  showed that both of these proteins interact with the Bag6 complex, but only Sec61 $\beta$  is primarily found bound to TRC40 (Supplementary Fig. 15). Given that the loading of tail-anchored proteins onto TRC40 depends on the Bag6 complex<sup>4</sup>, these data suggest that the Bag6 complex is acting as a triage



**Figure 2 | BAG6 interacts with MLPs through hydrophobic domains.** **a**, PrP translated in RRL or Fr-RRL, with or without 10  $\mu$ M ubiquitin (Ub), was analysed directly (left) or after anti-ubiquitin antibody immunoprecipitation (IP) (right) by using SDS-PAGE and autoradiography. **b**, PrP translated in Fr-RRL was ubiquitinated when UBC5 (E2; 250 nM) was included co-translationally (co) or post-translationally (post). Total synthesis (bottom) and ubiquitinated products (top) are shown. **c**, PrP was immunoaffinity purified under native conditions and incubated with the indicated components (cyt, cytosol; E1 enzyme, 100 nM; E2 enzyme, UBC5, 250 nM). All reactions contained His-ubiquitin and ATP. Purified ubiquitinated products are shown. **d**, PrP translated in Fr-RRL was separated into ten fractions in a 5–25% sucrose gradient. The fractions were subjected to chemical crosslinking (bottom) or ubiquitination assays (top). Asterisks indicate crosslinks. Histogram bars indicate the amount of ubiquitinated product in each fraction. The ~150-kDa crosslinking partner (x p150) is indicated. **e**, Crosslinking reactions (XL) of *in vitro*-synthesized PrP or PrP deletion constructs were analysed directly or after immunoprecipitation with anti-BAG6 or control (cont) antibodies. The crosslink to BAG6 (x BAG6) is indicated.



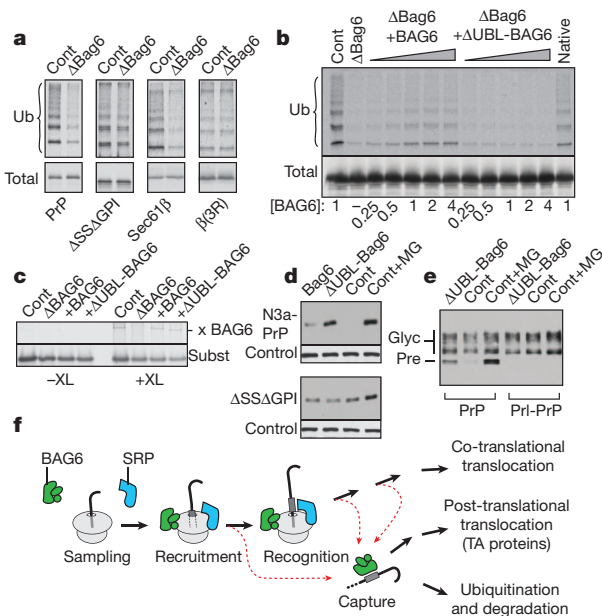
**Figure 3 | BAG6 captures MLPs released from the ribosome.** **a**, Diagram of constructs derived from Sec61 $\beta$ , with transmembrane domains shown as grey boxes and hydrophilic changes in white boxes. **b**, RNCs of  $\beta$ -CFP with the TMD outside the ribosome were subjected to crosslinking (XL) before or after release by puromycin (puro) and were analysed directly (bottom) or after immunoprecipitation (IP) with anti-BAG6 antibody (top) or anti-SRP54 antibody (centre). The results are also illustrated diagrammatically: Bag6 complex, green; SRP, blue; and ribosome, pale grey. **c**, The assays were as described in **b** but using TR- $\beta$  (top) and RT- $\beta$  (bottom). **d**, The indicated constructs were translated *in vitro*, immunoaffinity purified through their N terminus, and immunoblotted with anti-TRC40 antibody or anti-UBL4A antibody (the latter to detect the Bag6 complex). The autoradiograph shows equal recovery of the translated substrates.



factor: that is, it captures a relatively broad range of membrane proteins after their ribosomal release but transfers only a subset of them (namely, tail-anchored proteins) to TRC40 for post-translational membrane insertion. The remainder seem to be targeted for ubiquitination because of their persistent interaction with BAG6.

To examine this hypothesis, we immunodepleted the Bag6 complex from RRL (Supplementary Fig. 16) and found that the ubiquitination of several MLPs was reduced (Fig. 4a and Supplementary Fig. 17). By contrast, the control protein GFP was not ubiquitinated in RRL but became a substrate when it was attached to either a ubiquitin molecule or any of several hydrophobic ER-targeting domains (Supplementary Fig. 18). Only the hydrophobically modified GFP proteins were BAG6 dependent in their ubiquitination, consistent with their interaction with BAG6 by crosslinking analysis (Supplementary Fig. 13). Conversely,  $\Delta$ SS $\Delta$ GPI-PrP, which does not interact with BAG6 (Fig. 2e), was ubiquitinated (albeit slowly and less efficiently) in a BAG6-independent manner (Fig. 4a). Disrupting the TMD of Sec61 $\beta$  with three arginine residues (denoted  $\beta$ (3R)), which disrupts BAG6 interaction<sup>4</sup>, also resulted in less ubiquitination, which was no longer BAG6 dependent (Fig. 4a). Thus, the Bag6 complex is not required for ubiquitination of all misfolded proteins but is especially important for the efficient ubiquitination of MLPs.

When recombinant BAG6 (Supplementary Fig. 16) was added to translation extracts that had been depleted of the Bag6 complex, the ubiquitination of a model MLP was restored (Fig. 4b), and the recombinant BAG6 interacted with this MLP in crosslinking assays (Fig. 4c).



**Figure 4 | Maximum ubiquitination of MLPs requires BAG6.** **a**, Various constructs (listed at bottom) were assayed for ubiquitination in lysates containing Bag6 complex (control, cont) or lacking Bag6 complex ( $\Delta$ Bag6). The gels for assessing ubiquitination for the  $\Delta$ SS $\Delta$ GPI and  $\beta$ (3R) constructs were exposed about threefold longer than those for PrP and Sec61 $\beta$ . **b**, Bag6-complex-depleted lysates ( $\Delta$ Bag6) were replenished with increasing amounts (wedges) of recombinant BAG6 (Supplementary Fig. 16),  $\Delta$ UBL-BAG6 or native Bag6 complex and then analysed for the ubiquitination of TR- $\beta$ . Relative BAG6 levels are indicated (listed at bottom). **c**, TR- $\beta$  interacts with recombinant BAG6 and  $\Delta$ UBL-BAG6 by crosslinking (XL). Subst, substrate; x BAG6, crosslink to BAG6. **d**, The indicated PrP constructs (N3a-PrP and  $\Delta$ SS $\Delta$ GPI) were co-transfected with Bag6 complex,  $\Delta$ UBL-Bag6 complex or control plasmid (cont) (Supplementary Fig. 20), and PrP was detected by immunoblotting. One sample was treated with the proteasome inhibitor MG132 (MG) for 4 h. A loading control (control) is also shown. **e**, Effect of the  $\Delta$ UBL-Bag6 complex on wild-type PrP and Prl-PrP. Unglycosylated precursor PrP (pre) is preferentially stabilized by either overexpression of the  $\Delta$ UBL-Bag6 complex or inhibition of the proteasome. **f**, The model we propose is that the Bag6 complex captures ribosomally released hydrophobic proteins (red arrows) and triages them between post-translational targeting (for tail-anchored (TA) proteins) and ubiquitination.

BAG6 lacking its N-terminal UBL domain ( $\Delta$ UBL-BAG6) was inactive in restoring ubiquitination (Fig. 4b) despite interacting normally with substrate (Fig. 4c). This finding suggested that BAG6 may recruit the ubiquitination machinery to substrates through its UBL domain. To test this, Flag-tagged recombinant BAG6 or  $\Delta$ UBL-BAG6 was added to the Fr-RRL translation system lacking the E2 enzyme UBCH5 (Supplementary Fig. 7). BAG6–substrate complexes were immunopurified through the Flag tag and incubated with purified E1 ubiquitin-activating enzyme, E2 enzyme, ubiquitin and ATP. Substrate ubiquitination was observed with BAG6 but not  $\Delta$ UBL-BAG6, verifying that the UBL domain recruits the ubiquitination machinery to the substrate (Supplementary Fig. 19). Indeed, BAG6 has been observed to interact with E3 ubiquitin ligases through its UBL domain<sup>21,28</sup>.

In Fig. 4b, c, the data indicated that  $\Delta$ UBL-BAG6 should act as a dominant negative and partly stabilize BAG6 substrates, thereby providing a selective tool for *in vivo* analysis. We therefore overexpressed the Bag6 complex or the  $\Delta$ UBL-Bag6 complex (by about twofold) (Supplementary Fig. 20) in cultured cells and assessed the levels of a co-expressed MLP substrate. A translocation-impaired signal-sequence mutant of PrP (termed N3a-PrP)<sup>5</sup> was stabilized by the  $\Delta$ UBL-Bag6 complex but almost unaffected by the wild-type Bag6 complex (Fig. 4d). Importantly,  $\Delta$ SS $\Delta$ GPI-PrP, which does not interact with BAG6 (Fig. 2e), was unaffected by either Bag6 complex or  $\Delta$ UBL-Bag6 complex overexpression (Fig. 4d) and showed higher steady-state levels than N3a-PrP (data not shown). This finding suggests that degradation is occurring by a different quality control pathway, consistent with the failure of  $\Delta$ SS $\Delta$ GPI-PrP to be recognized as an MLP (Fig. 2e).

Wild-type PrP, the translocation of which is slightly inefficient *in vivo*<sup>2,3,6,8–10</sup>, showed preferential stabilization of a non-glycosylated species when co-overexpressed with  $\Delta$ UBL-Bag6 complexes (Fig. 4e and Supplementary Fig. 21). This species was stabilized by proteasome inhibition and had been shown in earlier studies to be a non-translocated PrP precursor<sup>2,3,9,10</sup>. Replacing the slightly inefficient PrP signal sequence with the efficient signal from preprolactin (Prl-PrP) precluded the generation of non-glycosylated PrP with either proteasome inhibition or  $\Delta$ UBL-Bag6 complex overexpression (Fig. 4e). Although the extent of stabilization seems modest, it is comparable to that seen after 2 h proteasome inhibition (Supplementary Fig. 21). Partial knockdown of BAG6 with a short hairpin RNA (shRNA) similarly stabilized a non-glycosylated species of PrP (Supplementary Fig. 22). Thus, MLPs are not only generated *in vivo*<sup>2,3,6,8–10</sup>, but also require functional BAG6 for maximally efficient degradation.

Our results reveal a pathway for MLP degradation and identify an unexpectedly close link with protein targeting (Fig. 4f). Ribosomes synthesizing nascent membrane proteins can recruit both the SRP and Bag6 complex on entry of the first hydrophobic segment into the ribosomal tunnel<sup>4,18</sup>. This is a potential targeting complex for the ER membrane in both the co-translational and post-translational membrane-protein insertion pathways. We now find that such ribosomes are also potential degradation complexes because the first component of this degradation pathway is already poised to act in the event of failed targeting or inappropriate release from the ribosome. BAG6 therefore imposes a degradative fate on membrane proteins that can be avoided only by productive targeting.

Because membrane proteins would never fold in the cytosol, their direct degradation by a specialized pathway may be important to avoid unnecessarily occupying essential cellular folding pathways, particularly under conditions of stress. MLPs are distinguished from nascent cytosolic proteins by relatively long linear hydrophobic stretches, a feature that is important for BAG6 recognition. Indeed, mutagenesis shows that even modest reductions of TMD hydrophobicity sharply curtail BAG6 interaction<sup>4</sup>. This specificity distinguishes BAG6 from more general chaperones such as heat-shock protein 70 (HSP70), the substrate-binding pocket of which seems more suited to the shorter, moderately hydrophobic segments that typify nascent cytosolic proteins. This differential specificity probably explains how MLPs are triaged differently from other

potential substrates of cytosolic quality control<sup>15,16,22–28</sup>. These pathways could intersect or cooperate in as yet undefined ways given that BAG6 and HSP70 have been observed to co-immunoprecipitate<sup>26</sup>.

In addition to this role in degradation, the Bag6 complex also facilitates the loading of tail-anchored proteins onto TRC40 for post-translational insertion into the ER<sup>4</sup>. As expected, tail-anchored proteins were also ubiquitinated by way of BAG6 in the absence of, or saturation of, TRC40 (Supplementary Fig. 23). Thus, substrates of both the co-translational and post-translational targeting pathways are ubiquitinated in a BAG6-dependent manner when targeting fails. After ubiquitination, BAG6 might chaperone its polyubiquitinated substrates to the proteasome, a function that was recently proposed on the basis of the co-immunoprecipitation of BAG6 with polyubiquitinated proteins<sup>26</sup>. The Bag6 complex is therefore a multi-purpose triage factor for chaperoning especially hydrophobic proteins through the aqueous cytosol. This view conceptually links its roles in tail-anchored protein targeting<sup>4,17</sup>, in the MLP pathway (in this study), as a chaperone for newly dislocated proteins during ER-associated protein degradation<sup>27,28</sup> and in the delivery of terminally misfolded proteins to the proteasome<sup>26</sup>.

## METHODS SUMMARY

**Reagents and standard methods.** The plasmids and antibodies used and the assays carried out (*in vitro* translation assays, sucrose gradient separations, chemical crosslinking analyses, immunoprecipitation assays and immunodepletion assays) were as previously described<sup>2–8,14,20,29,30</sup>. Pull-down assays with Co<sup>2+</sup> immobilized on chelating sepharose were performed on samples that had been denatured in boiling 1% SDS and then diluted tenfold in 4 °C pull-down buffer: 0.5% Triton X-100, 25 mM HEPES, 100 mM NaCl and 10 mM imidazole. Culture, transfection and immunoblotting analysis of N2a cells (dominant-negative inhibition experiments) and HeLa cells (for shRNA experiments) were carried out as previously described<sup>2,3</sup>. Full-length BAG6 (or ΔUBL-BAG6, which lacks residues 15–89) tagged at the C terminus with a Flag epitope was overexpressed after transient transfection of HEK-293T cells and then purified with anti-Flag resin under high salt (400 mM potassium acetate) conditions.

**Modified translation extracts.** Fr-RRL contained native ribosomes (isolated from RRL) mixed with a diethylaminoethyl (DEAE) sepharose ion-exchange chromatography elution fraction prepared from ribosome-free RRL (Supplementary Fig. 7). Fr-RRL was adjusted to the following final conditions for translation: 72 mM potassium acetate, 2.5 mM magnesium acetate, 10 mM HEPES, pH 7.4, 2 mM dithiothreitol (DTT), 0.2 mg ml<sup>-1</sup> liver transfer RNA, 1 mM ATP, 1 mM GTP, 12 mM creatine phosphate, 40 μg ml<sup>-1</sup> creatine kinase, 40 μM each amino acid (except methionine) and 1 μCi μl<sup>-1</sup> [<sup>35</sup>S]methionine.

**Ubiquitination assays.** For full-length proteins, translations containing 10 μM His-tagged ubiquitin were carried out for 1 h at 32 °C. In Fr-RRL, post-translational ubiquitination was initiated by adding E2 enzyme to a final concentration of 250 nM and incubating for 1 h at 32 °C. For RNCs, samples were supplemented with E1 enzyme (85 nM), E2 enzyme (usually 250 nM or 500 nM), cytosol (RRL or Fr-RRL), 10 μM His-ubiquitin, an ATP-regenerating system (1 mM ATP, 10 mM creatine phosphate and 40 μg ml<sup>-1</sup> creatine kinase) and 1 mM puromycin. The reaction conditions were 100 mM potassium acetate, 50 mM HEPES, pH 7.4, 5 mM MgCl<sub>2</sub> and 1 mM DTT. Incubations were carried out for 1 h at 32 °C. On-bead ubiquitination of affinity-purified products was carried out under the same conditions, except without the inclusion of puromycin.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 8 December 2010; accepted 6 May 2011.**

**Published online 10 July 2011.**

1. Cross, B. C., Sinning, I., Lührink, J. & High, S. Delivering proteins for export from the cytosol. *Nature Rev. Mol. Cell Biol.* **10**, 255–264 (2009).
2. Rane, N. S., Yonkovich, J. L. & Hegde, R. S. Protection from cytosolic prion protein toxicity by modulation of protein translocation. *EMBO J.* **23**, 4550–4559 (2004).
3. Kang, S. W. *et al.* Substrate-specific translocational attenuation during ER stress defines a pre-emptive quality control pathway. *Cell* **127**, 999–1013 (2006).
4. Mariappan, M. *et al.* A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* **466**, 1120–1124 (2010).
5. Kim, S. J., Mitra, D., Salerno, J. R. & Hegde, R. S. Signal sequences control gating of the protein translocation channel in a substrate-specific manner. *Dev. Cell* **2**, 207–217 (2002).

6. Levine, C. G., Mitra, D., Sharma, A., Smith, C. L. & Hegde, R. S. The efficiency of protein compartmentalization into the secretory pathway. *Mol. Biol. Cell* **16**, 279–291 (2005).
7. Kim, S. J. & Hegde, R. S. Cotranslational partitioning of nascent prion protein into multiple populations at the translocation channel. *Mol. Biol. Cell* **13**, 3775–3786 (2002).
8. Rane, N. S., Chakrabarti, O., Feigenbaum, L. & Hegde, R. S. Signal sequence insufficiency contributes to neurodegeneration caused by transmembrane prion protein. *J. Cell Biol.* **188**, 515–526 (2010).
9. Orsi, A., Fioriti, L., Chiesa, R. & Sitia, R. Conditions of endoplasmic reticulum stress favor the accumulation of cytosolic prion protein. *J. Biol. Chem.* **281**, 30431–30438 (2006).
10. Drisaldi, B. *et al.* Mutant PrP is delayed in its exit from the endoplasmic reticulum, but neither wild-type nor mutant PrP undergoes retrotranslocation prior to proteasomal degradation. *J. Biol. Chem.* **278**, 21732–21743 (2003).
11. Ma, J. & Lindquist, S. Conversion of PrP to a self-perpetuating PrPSc-like conformation in the cytosol. *Science* **298**, 1785–1788 (2002).
12. Chakrabarti, O. & Hegde, R. S. Functional depletion of mahogunin by cytosolically exposed prion protein contributes to neurodegeneration. *Cell* **137**, 1136–1147 (2009).
13. Ma, J., Wollmann, R. & Lindquist, S. Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* **298**, 1781–1785 (2002).
14. Rane, N. S., Kang, S. W., Chakrabarti, O., Feigenbaum, L. & Hegde, R. S. Reduced translocation of nascent prion protein during ER stress contributes to neurodegeneration. *Dev. Cell* **15**, 359–370 (2008).
15. Buchberger, A., Bukau, B. & Sommer, T. Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* **40**, 238–252 (2010).
16. McDonough, H. & Patterson, C. CHIP: a link between the chaperone and proteasome systems. *Cell Stress Chaperones* **8**, 303–308 (2003).
17. Leznicki, P., Clancy, A., Schwappach, B. & High, S. Bat3 promotes the membrane integration of tail-anchored proteins. *J. Cell Sci.* **123**, 2170–2178 (2010).
18. Berndt, U., Oellerer, S., Zhang, Y., Johnson, A. E. & Rospert, S. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc. Natl Acad. Sci. USA* **106**, 1398–1403 (2009).
19. Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. The signal recognition particle. *Annu. Rev. Biochem.* **70**, 755–775 (2001).
20. Stefanovic, S. & Hegde, R. S. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* **128**, 1147–1159 (2007).
21. Lehner, B. *et al.* Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* **83**, 153–167 (2004).
22. Park, S. H. *et al.* The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin-proteasome system. *Mol. Biol. Cell* **18**, 153–165 (2007).
23. Eisele, F. & Wolf, D. H. Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. *FEBS Lett.* **582**, 4143–4146 (2008).
24. Heck, J. W., Cheung, S. K. & Hampton, R. Y. Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proc. Natl Acad. Sci. USA* **107**, 1106–1111 (2010).
25. Nillegoda, N. B. *et al.* Ubr1 and Ubr2 function in a quality control pathway for degradation of unfolded cytosolic proteins. *Mol. Biol. Cell* **21**, 2102–2116 (2010).
26. Minami, R. *et al.* BAG-6 is essential for selective elimination of defective proteasomal substrates. *J. Cell Biol.* **190**, 637–650 (2010).
27. Ernst, R. *et al.* Enzymatic blockade of the ubiquitin-proteasome pathway. *PLoS Biol.* **8**, e1000605 (2011).
28. Wang, Q. *et al.* A chaperone holdase maintains polypeptides in soluble states for proteasome degradation. *Mol. Cell* doi:10.1016/j.molcel.2011.05.010 (in the press).
29. Garrison, J. L., Kunkel, E. J., Hegde, R. S. & Taunton, J. A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum. *Nature* **436**, 285–289 (2005).
30. Sharma, A., Mariappan, M., Appathurai, S. & Hegde, R. S. *In vitro* dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol.* **619**, 339–363 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to E. Whiteman and X. Li for carrying out the initial experiments for parts of this project, S.W. Kang, S. Shao, and Z. Zhang for discussions, P. Sengupta, J. Magadán, and C. Ott for constructs, J. Taunton and J. Garrison for cotransin, S. Shao for comments on the manuscript, and Y. Ye for discussions and sharing results before publication. This work was supported by the Intramural Research Program of the National Institutes of Health (R.S.H.) and a postdoctoral fellowship from The Wenner-Gren Foundations (T.H.).

**Author Contributions** T.H. performed most of the experiments, with contributions from A.S. (ubiquitination assays in modified lysates), M.M. (defining the substrate specificity of BAG6), H.D.E. (characterizing the Fr-RRL system), E.G. (BAG6 crosslinking analysis) and R.S.H. (*in vivo* studies). R.S.H. conceived the project, guided the experiments and wrote the paper with input from all of the authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.S.H. ([hegde.science@gmail.com](mailto:hegde.science@gmail.com)).



## METHODS

**Plasmids and antibodies.** The SP64 vector-based constructs encoding bovine preprolactin, PrP, ΔSS-PrP (lacking residues 2–22), ΔSSΔGPI-PrP (additionally lacking residues 232–254) and HA-tagged PrP (with the epitope inserted at codon 50) have been characterized previously<sup>5,29–32</sup>. PrP-PrP and NPY-PrP encode versions in which the N-terminal signal sequence (residues 1–22) of PrP was replaced<sup>5</sup> with that of either bovine preprolactin or human neuropeptide Y (NPY). N3a-PrP contains a mutated signal sequence (WL was replaced with DD at residues 7 and 8) that is translocation deficient<sup>5</sup>. The lysine-free version of PrP was provided by C. Ott and made by standard mutagenesis methods. Wild-type Sec61β (appended at the C terminus with an epitope recognized by the 3F4 antibody), Sec61β(3R), Sec61β-CFP and CFP–Sec61β have been described previously<sup>4,20</sup>. Sec61β-TR (referred to as TR-β in the text and figures) contains the TMD of the human transferrin receptor (IAVIVFLLIGFMIGYLG) at codon 50 in the cytosolic domain of Sec61β<sup>4</sup>. This positions the TMD of TR outside the ribosomal tunnel when the Sec61β TMD is inside the tunnel<sup>4</sup>. RT-β contains an irrelevant hydrophilic sequence (YKYPIMNPIKKKTITAI) at the same position<sup>4</sup>. GFP, SS/GPI-GFP (containing the N-terminal signal sequence of bovine preprolactin and the C-terminal GPI anchoring sequence of PrP), ManII-GFP (containing the N-terminal type II signal anchor domain of Golgi α-mannosidase II) and SiT-GFP (containing the type II signal anchor domain of sialyl transferase) have been described previously<sup>32–34</sup>. The plasmid encoding Vpu (a type-I-signal-anchored membrane protein from HIV-1) was obtained from J. Bonifacino and J. Magadán<sup>35</sup>. An expression plasmid for bovine rhodopsin has been characterized<sup>29</sup>. For translations of full-length products, the open reading frames were PCR amplified using a forward 5′ primer annealing to or encoding an SP6 or T7 promoter, and a reverse primer in the 3′ untranslated region at least 100 nucleotides beyond the stop codon. For RNCs, the reverse primer annealed in the coding region and lacked a stop codon. PrP and Vpu RNCs included the entire open reading frame except for the stop codon. The RNCs of β-CFP encoded 46 residues beyond the TMD such that this domain would fully emerge from the ribosome. Similarly, the RNCs of TR-β and RT-β encoded up to and including the TMD of Sec61β such that the TR and RT sequences emerge from the ribosome. Genetic constructs encoding BAG6-Flag and ΔUBL-BAG6-Flag (lacking residues 15–89 of BAG6)—both encoding human BAG6 containing a C-terminal Flag epitope—were subcloned into a mammalian expression vector by using standard methods. Expression vectors for human TRC35 and UBL4A containing C-terminal Flag tags were obtained from OriGene. Expression vectors for shRNAs directed against human BAG6 were from OriGene. The target sequences were TGACGGCT CTGCTGTGGATGTTACATCA and CAGCTATGTCATGGTTGGAACCT TCAATC. The irrelevant sequence used as a control was GCACTACCAG AGCTAACTCAGATAGTACT. Antibodies specific for BAG6, TRC40, TRC35, UBL4A and Sec61β have been described previously<sup>4,36</sup>. Anti-SRP54 (BD Biosciences), anti-ubiquitin (BIOMOL), and 3F4 anti-PrP monoclonal antibodies (Signet) were purchased.

**In vitro translation.** *In vitro* transcription and translation in RRL was carried out with minor modifications to published procedures<sup>30</sup>. The most notable change was the inclusion in most experiments of 10 μM His-tagged ubiquitin (Boston Biochem) to facilitate the subsequent isolation of ubiquitinated products. Preliminary experiments showed that, at this concentration, endogenous ubiquitin was more than 90% competed out, resulting in few or no untagged ubiquitinated products. Translation times, unless otherwise indicated, were 1 h at 32 °C. Shorter times for tail-anchored proteins (as used in our earlier studies) resulted in very little ubiquitination<sup>4,20</sup>, presumably because saturation of TRC40 is required before substrates occupy the Bag6 complex<sup>4</sup>. To generate RNCs, the translation times were typically reduced to 30 min to minimize spontaneous release or hydrolysis of the tRNA. Translocation assays into rough microsomes<sup>5</sup>, inhibition by cotransin<sup>29</sup> and inactivation with NEM<sup>37</sup> treatment were carried out as previously described. For direct analysis or downstream immunoprecipitation, translation reactions were stopped, and the proteins were denatured using 1% SDS and heating to 100 °C. For other applications requiring native complexes (for example, crosslinking, affinity purification or downstream assays), samples were placed on ice, and subsequent manipulations were performed at 0–4 °C.

**Sucrose gradient separation and crosslinking.** To generate RNCs, translation reactions (typically 200 μl volume) were chilled on ice and immediately layered onto 2-ml 10–50% sucrose gradients in physiological salt buffer (PSB; 100 mM potassium acetate, 50 mM HEPES, pH 7.4, and 2 mM magnesium acetate). Centrifugation was carried out for 1 h at 55,000 r.p.m. at 4 °C in a TLS-55 rotor (Beckman), after which 200 μl fractions were removed from the top. The peak ribosomal fractions (6 and 7) were pooled and used as the RNCs. These were used immediately or flash frozen in liquid nitrogen for later use in RNC crosslinking or ubiquitination experiments. Chemical crosslinking experiments were essentially carried out as described previously<sup>4,20</sup>. Chilled translation reactions were layered

onto 2-ml 5–25% sucrose gradients in PSB and centrifuged for 5 h at 55,000 r.p.m. at 4 °C in a TLS-55 rotor, after which 200 μl fractions were removed from the top. Crosslinking experiments used 250 μM BMH, except for in experiments to detect SRP interaction, which used 200 μM DSS. Reactions were carried out for 30 min at either 0 °C (BMH) or 25 °C (DSS) and quenched with 25 mM 2-mercaptoethanol (BMH) or 100 mM Tris (DSS). The samples were subsequently denatured and subjected to direct analysis or immunoprecipitation as described below. Photocrosslinking was carried out by following published methods<sup>38</sup>, except that we used the Fr-RRL system for translation and benzophenone-modified lysyl-tRNA (tRNA Probes). The absence of endogenous charged tRNAs and haemoglobin increased photocrosslinker incorporation and photolysis, respectively. Photolysis was carried out for 15 min on ice, and the samples were analysed directly.

**Modified translation extracts.** Fr-RRL was typically prepared from 25 ml RRL (Green Hectares) that had first been treated with haemin and micrococcal nuclease. Its characterization will be described in a future publication, but its preparation is as follows. All procedures were carried out on ice or at 4 °C. The lysate was centrifuged at 100,000 r.p.m. for 40 min in a TLA100.4 rotor (Beckman). The supernatants were pooled, and the tubes rinsed (without disrupting the ribosomal pellet) with an equal volume of column buffer (20 mM Tris, pH 7.5, 20 mM KCl, 0.1 mM EDTA and 10% glycerol), which was added to the supernatant. The pellet was resuspended by dounce homogenization in ribosome wash buffer (RWB; 20 mM HEPES, pH 7.5, 100 mM potassium acetate, 1.5 mM magnesium acetate and 0.1 mM EDTA), layered onto a 1 M sucrose cushion in RWB, and re-isolated by centrifugation at 100,000 r.p.m. for 1 h in a TLA100.4 rotor. The final pellet was resuspended in one-tenth of the original lysate volume and defined as ‘native ribosomes’. The ribosome-free supernatant from above was applied to a 10 ml DEAE column at a flow rate of ~1 ml min<sup>-1</sup> and washed with column buffer until the red haemoglobin was removed (~50 ml). The elution was carried out in a single step with 50 ml column buffer containing 300 mM KCl. The eluate was adjusted slowly with solid ammonium sulphate to 75% saturation (at 4 °C) with constant stirring. After 1 h mixing, the precipitate was recovered by centrifugation at 15,000 r.p.m. in a JA-17 rotor (Beckman). The supernatant was discarded, and the pellet was dissolved in a minimal volume (~8 ml) of dialysis buffer (20 mM HEPES, pH 7.4, 100 mM potassium acetate, 1.5 mM magnesium acetate, 10% glycerol and 1 mM DTT). This solution was dialysed against two changes of dialysis buffer overnight, recovered, adjusted to 10–12 ml (that is, twice the original concentration) and flash frozen in liquid nitrogen. To make a translation-competent Fr-RRL, the native ribosomes and dialysed DEAE eluate were adjusted to 72 mM potassium acetate, 2.5 mM magnesium acetate, 10 mM HEPES, pH 7.4, 2 mM DTT, 0.2 mg ml<sup>-1</sup> liver tRNA, 1 mM ATP, 1 mM GTP, 12 mM creatine phosphate, 40 μg ml<sup>-1</sup> creatine kinase, 40 μM each amino acid (except for methionine) and 1 μCi μl<sup>-1</sup> [<sup>35</sup>S]methionine. The concentration of ribosomes and lysate was the same as that for RRL. Immunodepletions of RRL were carried out as described previously<sup>4</sup>.

**Ubiquitination assays.** The human E1 enzyme and all mammalian E2 enzymes were obtained from Boston Biochem. For full-length proteins, translations containing 10 μM His-ubiquitin were carried out for 1 h at 32 °C. In Fr-RRL, post-translational ubiquitination was initiated by adding E2 enzyme to a final concentration of 250 nM and further incubating for 1 h. For RNCs, samples were supplemented (as indicated in the figures) with E1 enzyme (85 nM), E2 enzyme (usually 250 or 500 nM), cytosol (RRL or Fr-RRL, at the same concentration as in the translations), 10 μM His-ubiquitin, an ATP-regenerating system (1 mM ATP, 10 mM creatine phosphate and 40 μg ml<sup>-1</sup> creatine kinase) and 1 mM puromycin. Reaction conditions were 100 mM potassium acetate, 50 mM HEPES, pH 7.4, 5 mM MgCl<sub>2</sub> and 1 mM DTT. Incubation was for 1 h at 32 °C. On-bead ubiquitination of affinity-purified products was carried out under the same conditions, except for without puromycin. To prepare the affinity-purified substrate, translation reactions in Fr-RRL were chilled on ice, diluted to 1 ml in PSB and incubated with immobilized antibodies against the HA epitope (for PrP-HA and Vpu-HA) or Sec61β. In Supplementary Fig. 9, the translation reactions were supplemented with Flag-tagged BAG6 or ΔUBL-BAG6 (each added to twofold excess above endogenous BAG6 levels), and anti-Flag beads (Sigma) were used for the pull-down. After 1 h, the resin was washed five times in PSB, and the residual buffer was carefully removed before adding the ubiquitination components as above. The reaction was incubated with constant low-level shaking (in a Thermomixer, Eppendorf) at 32 °C for 1 h. SDS (1%) was added directly to the reactions, which were analysed directly and after ubiquitin pull-downs.

**Cell culture studies.** Culture, transfection and immunoblotting analysis of N2a cells (dominant-negative inhibition experiments) and HeLa cells (for shRNA experiments) were carried out as described previously<sup>2,3</sup>. Cells were seeded in 24-well dishes the day before transfection. For the dominant-negative experiments, the plasmids were mixed in the ratios indicated in Supplementary Fig. 20 and transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. At 24 h after transfection, the cells were harvested in 1% SDS;



the DNA was sheared by vortexing and boiling; and the total sample was analysed by SDS-PAGE and immunoblotting. For shRNA experiments, each well received a mixture of 550 ng shRNA plasmid, 200 ng PrP expression plasmid and 50 ng CFP expression plasmid. Transfection was effected with Lipofectamine 2000. Examination of CFP fluorescence verified at least 50% transfection efficiency. The cells were cultured for ~100 h before collection and analysis by immunoblotting.

**BAG6 purification.** Full-length BAG6 or  $\Delta$ UBL-BAG6 tagged at the C terminus with a Flag epitope was overexpressed by transient transfection of HEK-293T cells. TransIT reagent (Mirus) was used. After 3 days of expression, the cells were collected in 50 mM HEPES, pH 7.4, 150 mM potassium acetate, 5 mM magnesium acetate and 1% deoxy Big CHAP. The soluble extract was incubated with immobilized anti-Flag antibodies (Sigma) with constant mixing, and the resin was washed four times with high salt lysis buffer containing 400 mM potassium acetate and then twice with detergent-free lysis buffer containing 230 mM potassium acetate. Elution was carried out with 1 mg ml<sup>-1</sup> competing peptide at room temperature. The final protein was checked by using colloidal Coomassie blue (Supplementary Fig. 16), and its concentration relative to that in RRL was determined by immunoblotting of serial dilutions. Blotting also confirmed the lack of TRC35 and UBL4A in BAG6 prepared by this method.

**Miscellaneous biochemistry.** Immunoprecipitation assays were carried out as described previously<sup>5,36</sup>. Pull-down assays with Co<sup>2+</sup> immobilized on chelating sepharose were performed on samples denatured in boiling 1% SDS and then diluted tenfold in cold (4 °C) 0.5% Triton X-100, 25 mM HEPES, 100 mM NaCl and 10 mM imidazole. The complete denaturation step is essential for samples containing RRL because the haemoglobin is a strong Co<sup>2+</sup>-binding protein in its

native state. Typically, 10  $\mu$ l packed resin was used per sample, and after incubation for 1–2 h at 4 °C, the resin was washed three times in the above buffer and eluted in SDS-PAGE sample buffer containing 20 mM EDTA. SDS-PAGE was carried out using 8.5% or 12% tricine gels. Figures were prepared using the programs Photoshop and Illustrator (Adobe).

31. Emerman, A. B., Zhang, Z. R., Chakrabarti, O. & Hegde, R. S. Compartment-restricted biotinylation reveals novel features of prion protein metabolism *in vivo*. *Mol. Biol. Cell* **21**, 4325–4337 (2010).
32. Ashok, A. & Hegde, R. S. Retrotranslocation of prion proteins from the endoplasmic reticulum by preventing GPI signal transamidation. *Mol. Biol. Cell* **19**, 3463–3476 (2008).
33. Cole, N. B. *et al.* Diffusional mobility of Golgi proteins in membranes of living cells. *Science* **273**, 797–801 (1996).
34. Wu, M. M. *et al.* Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* **7**, 197–209 (2000).
35. Magadán, J. G. *et al.* Multilayered mechanism of CD4 downregulation by HIV-1 Vpu involving distinct ER retention and ERAD targeting steps. *PLoS Pathogens* **6**, e1000869 (2010).
36. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* **160**, 529–539 (2003).
37. Gilmore, R., Blobel, G. & Walter, P. Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *J. Cell Biol.* **95**, 463–469 (1982).
38. Krieg, U. C., Walter, P. & Johnson, A. E. Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. *Proc. Natl Acad. Sci. USA* **83**, 8604–8608 (1986).

# The ELF4–ELF3–LUX complex links the circadian clock to diurnal control of hypocotyl growth

Dmitri A. Nusinow<sup>1,2</sup>, Anne Helfer<sup>1,2</sup>, Elizabeth E. Hamilton<sup>1,2</sup>, Jasmine J. King<sup>1,2</sup>, Takato Imaizumi<sup>1†</sup>, Thomas F. Schultze<sup>1†</sup>, Eva M. Farré<sup>1‡</sup> & Steve A. Kay<sup>1,2</sup>

The circadian clock is required for adaptive responses to daily and seasonal changes in environmental conditions<sup>1–3</sup>. Light and the circadian clock interact to consolidate the phase of hypocotyl cell elongation to peak at dawn under diurnal cycles in *Arabidopsis thaliana*<sup>4–7</sup>. Here we identify a protein complex (called the evening complex)—composed of the proteins encoded by *EARLY FLOWERING 3* (*ELF3*), *ELF4* and the transcription-factor-encoding gene *LUX ARRHYTHMO* (*LUX*; also known as *PHYTOCLOCK 1*)—that directly regulates plant growth<sup>8–12</sup>. *ELF3* is both necessary and sufficient to form a complex between *ELF4* and *LUX*, and the complex is diurnally regulated, peaking at dusk. *ELF3*, *ELF4* and *LUX* are required for the proper expression of the growth-promoting transcription factors encoded by *PHYTOCHROME INTERACTING FACTOR 4* (*PIF4*) and *PIF5* (also known as *PHYTOCHROME INTERACTING FACTOR 3-LIKE 6*) under diurnal conditions<sup>4,6,13</sup>. *LUX* targets the complex to the promoters of *PIF4* and *PIF5* *in vivo*. Mutations in *PIF4* and/or *PIF5* are epistatic to the loss of the *ELF4*–*ELF3*–*LUX* complex, suggesting that regulation of *PIF4* and *PIF5* is a crucial function of the complex. Therefore, the evening complex underlies the molecular basis for circadian gating of hypocotyl growth in the early evening.

The circadian clock is an endogenous molecular oscillator with a period of ~24 h that is almost ubiquitous<sup>1</sup>. In plants, multiple interlocking transcriptional feedback loops contribute to the robust architecture of this oscillator network<sup>3</sup>. The clock functions to enable anticipation of diurnal, rhythmic environmental changes, allowing optimal phasing of molecular, physiological and behavioural responses to specific times of day<sup>2</sup>. Plant growth is a physiological response that is controlled by both the clock and the changes in light conditions; under diurnal growth conditions, maximal plant growth occurs at the end of night<sup>4–7</sup>.

*ELF3* and *ELF4* were first identified in genetic screens for photoperiodism mutants and were found to regulate circadian rhythms<sup>8–10,14</sup>. *ELF3* and *ELF4* encode plant-specific nuclear proteins with no known functional domains<sup>9,10,15,16</sup>. *LUX* is a single-MYB-domain-containing, SHAQYF-type GARP transcription factor that was identified in a genetic screen for long hypocotyl mutants and aberrant circadian-regulated gene expression<sup>11,12</sup>. The mutants *elf3*, *elf4* and *lux* share multiple phenotypes, including an arrhythmic circadian oscillator, abnormal hypocotyl growth in diurnal cycles, and early flowering<sup>4,8–12,14,15,17</sup>. *ELF3*, *ELF4* and *LUX* showed similar expression profiles in microarray experiments (Supplementary Fig. 1; DIURNAL database, <http://diurnal.cgrb.oregonstate.edu> refs 18, 19), and these expression profiles were confirmed by quantitative PCR with reverse transcription (RT–PCR) analysis under both diurnal and circadian conditions (Fig. 1a).

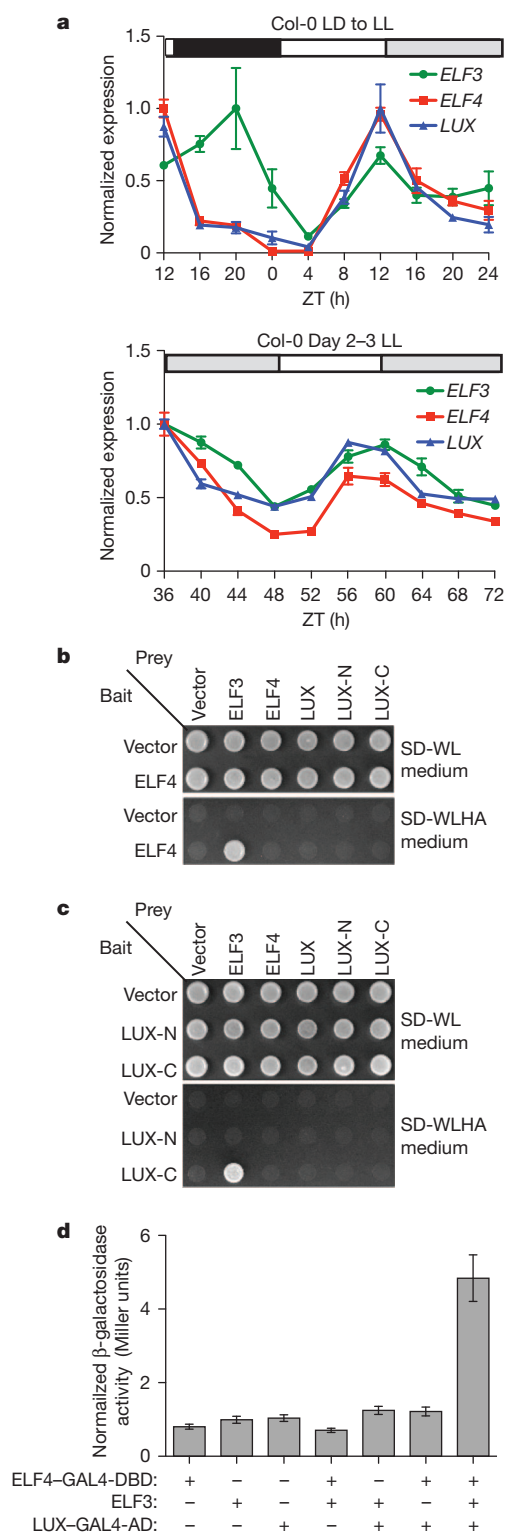
The similarities in expression patterns and phenotypes prompted us to test whether these proteins could interact. Using a yeast two-hybrid

assay, we found that *ELF4* interacted with *ELF3* (Fig. 1b). In addition, when *LUX* fragments were used as baits (full-length *LUX* showed auto-activation, data not shown), *ELF3* showed an interaction with *LUX*-C (amino acids 144–324), which contains the DNA-binding domain of *LUX*<sup>11,12,20</sup>, but not with *LUX*-N (amino acids 1–143) (Fig. 1c). *ELF4* did not interact with *LUX* or either *LUX* fragment (Fig. 1b, c). As *ELF3* could interact independently with either *ELF4* or *LUX*, we proposed that *ELF3* might form a complex between these two proteins. To test this, *ELF3* was used in a yeast three-hybrid system in combination with the fusion proteins *ELF4*–GAL4–DNA binding domain (GAL4–DBD) and/or *LUX*–GAL4–activation domain (GAL4–AD). Activation of the reporter was observed only when all three proteins were present, suggesting that *ELF3* was sufficient to bridge an interaction between *ELF4* and *LUX* (Fig. 1d).

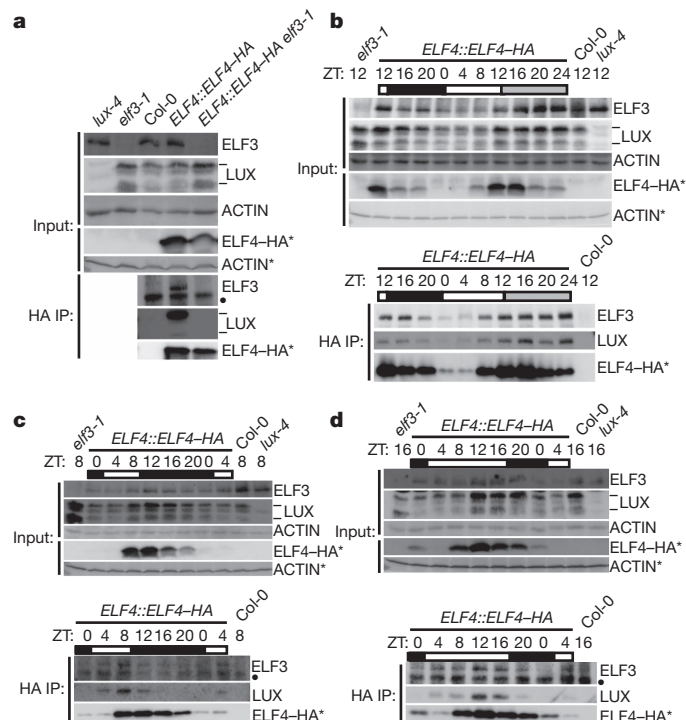
Next, we tested whether *ELF4*, *ELF3* and *LUX* interact *in vivo*. Antibodies were developed against *ELF3* and *LUX*, and an *ELF4::ELF4*–HA construct was introduced into the *elf4-2* mutant<sup>21</sup>. The encoded haemagglutinin (HA)-tagged *ELF4* protein is probably functional, because we identified transformants that rescued the hypocotyl length (Supplementary Fig. 2a) and circadian *CHLOROPHYLL A/B-BINDING PROTEIN::LUCIFERASE* (*CAB2::LUC*) rhythmicity, albeit with a shorter period than that of the wild type (Supplementary Fig. 2b–d). We then asked whether *ELF4*–HA could co-immunoprecipitate endogenous *ELF3* and/or *LUX* at Zeitgeber time 12 (ZT12) (Fig. 2a). We found that *ELF4*–HA could co-immunoprecipitate both *ELF3* and *LUX* (Fig. 2a and Supplementary Fig. 2f). The experiments in yeast suggested that *ELF3* bridges an interaction between *ELF4* and *LUX* and that *ELF3* would be necessary for the co-immunoprecipitation of *LUX* by *ELF4*–HA. To test this, we introduced *elf3-1* into the *ELF4::ELF4*–HA *elf4-2* transgenic line and immunoprecipitated *ELF4*–HA. Although similar amounts of *ELF4* and *LUX* were present in the extracts, *LUX* did not co-immunoprecipitate with *ELF4*–HA (Fig. 2a). These results show that *ELF3* is necessary for *in vivo* formation of the tripartite complex that includes *ELF4* and *LUX*. Furthermore, hypocotyl length in *elf3-1 elf4-3* and *elf3-1 lux-4* double mutants grown under a 12 h light and 12 h dark (12L:12D) cycle did not show additive effects over *elf3* (Supplementary Fig. 3). These results are consistent with the hypothesis that *ELF3*, *ELF4* and *LUX* function together as a complex to regulate common pathways.

Because *ELF4*, *ELF3* and *LUX* messenger RNA levels oscillate and peak with a similar phase, we analysed the dynamics of the protein levels under diurnal cycles. Tissue from the *ELF4::ELF4*–HA transgenic line was harvested every 4 h, starting at ZT12, under 12L:12D cycles, and then after transfer to constant light at ZT0 the following day. *ELF3*, *LUX* and *ELF4*–HA protein levels peaked at ZT12, declined during the night, reached a trough between ZT0 and ZT4 and then increased again (Fig. 2b and Supplementary Fig. 4). The levels of all three proteins remained elevated into the subjective dark period relative to their

<sup>1</sup>Section of Cell & Developmental Biology, Division of Biological Sciences, University of California San Diego 9500 Gilman Drive, La Jolla, California 92093-0130, USA. <sup>2</sup>Center for Chronobiology, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0116, USA. <sup>†</sup>Present addresses: Department of Biology, University of Washington, 24 Kincaid Hall, Box 351800, Seattle, Washington 98195-1800, USA (T.I.); Nicholas School of the Environment, Duke University Marine Laboratory, 135 Duke Marine Lab Road, Beaufort, North Carolina 28516, USA (T.F.S.); Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824-1312, USA (E.M.F.).



**Figure 1** | *ELF3*, *ELF4* and *LUX* are co-expressed, and *ELF3* directly interacts with both *ELF4* and *LUX* in yeast. **a**, Expression analysis by RT-PCR of *ELF3*, *ELF4* and *LUX* under diurnal or circadian conditions. Normalization is relative to the maximum. The rectangles above the graphs represent the light conditions during harvesting: black, lights off; white, lights on; and grey, lights on during subjective night. Error bars, s.e.m.;  $n = 3$ . LD, 12L:12D; LL, constant light. **b**, Yeast two-hybrid assay between *ELF4* and each of *ELF3*, *ELF4*, *LUX*, *LUX-N* and *LUX-C*. These experiments were repeated twice. **c**, Yeast two-hybrid assay carried out as for **b**, between *LUX* and each of *ELF3*, *ELF4* and *LUX*. **d**, Yeast three-hybrid assay assessing combinations of *ELF4*-GAL4-DBD, *LUX*-GAL4-AD and *ELF3*. Data are presented as fold induction over control vectors. Error bars, s.e.m.;  $n = 4$ .



**Figure 2** | *ELF3* bridges a diurnally regulated complex containing *ELF4* and *LUX* in vivo. **a**, *ELF3* is necessary for *ELF4* and *LUX* to co-precipitate in vivo. Immunoprecipitations (IPs) were performed on day 12 at ZT12 of a 12L:12D cycle. **b**, *ELF3*, *ELF4* and *LUX* oscillate over time and form a complex. **c**, **d**, EC formation in short and long days. Seedlings were grown under short-day or long-day photoperiods (8L:16D cycle or 16L:8D cycle, respectively) and harvested beginning at ZT0 on day 12. Experiments were performed three times with similar results. **a–d**, Rectangles indicate light conditions during harvesting as denoted in Fig. 1a. Upper panels show inputs, and lower panels HA IPs. —, each of the two *LUX* isoforms (low and high molecular weight); \*, *ELF4* separated in a 15% gel; •, background arising from HA-crosslinked beads.

respective time points in the dark, and the protein peak was shifted from ZT12 to the middle of the subjective night (Fig. 2b). Comparable results were observed for *ELF3* and *LUX* levels in wild-type seedlings (Supplementary Fig. 5). To assay time-dependent formation of the *ELF4*-*ELF3*-*LUX* complex (denoted the evening complex, EC), *ELF4*-HA was immunoprecipitated from the diurnal samples. The formation of the EC followed the same pattern as that of its composite parts, suggesting that these proteins would associate when present (Fig. 2b).

Photoperiodic control of flowering and growth is compromised in *elf3*, *elf4* and *lux* mutants<sup>4,8–12,14,15,17,22</sup>. To determine how *ELF4*, *ELF3* and *LUX* respond to altered photoperiods, we analysed the levels and formation of the EC in plants grown under short days (8L:16D) and long days (16L:8D). Peak levels of *ELF4*, *ELF3* and *LUX* followed their respective mRNA profiles under different photoperiods (Fig. 2c, d and Supplementary Fig. 4), similar to the findings of previous reports<sup>9,10,15,22</sup>. EC formation was also sensitive to photoperiod, peaking earlier in short days than in long days (Fig. 2c, d).

To investigate the molecular role of the EC, we focused on the diurnal hypocotyl growth phenotype shared by all mutants<sup>5,8–11,20</sup>. Previous work demonstrated that the basic helix–loop–helix transcription factors *PIF4* and *PIF5* are crucial for determining the hypocotyl elongation rate in seedlings, and that the genes encoding both factors act downstream of light- and clock-signalling pathways<sup>4,6,7,13,23</sup>. Expression of *PIF4* and *PIF5* was nearly antiphasic to that of the EC under different photocycles (Supplementary Fig. 6). This raised the possibility that the EC may be repressing the transcription of *PIF4* and *PIF5*, which is consistent with recent reports that *ELF3* and *LUX* act as

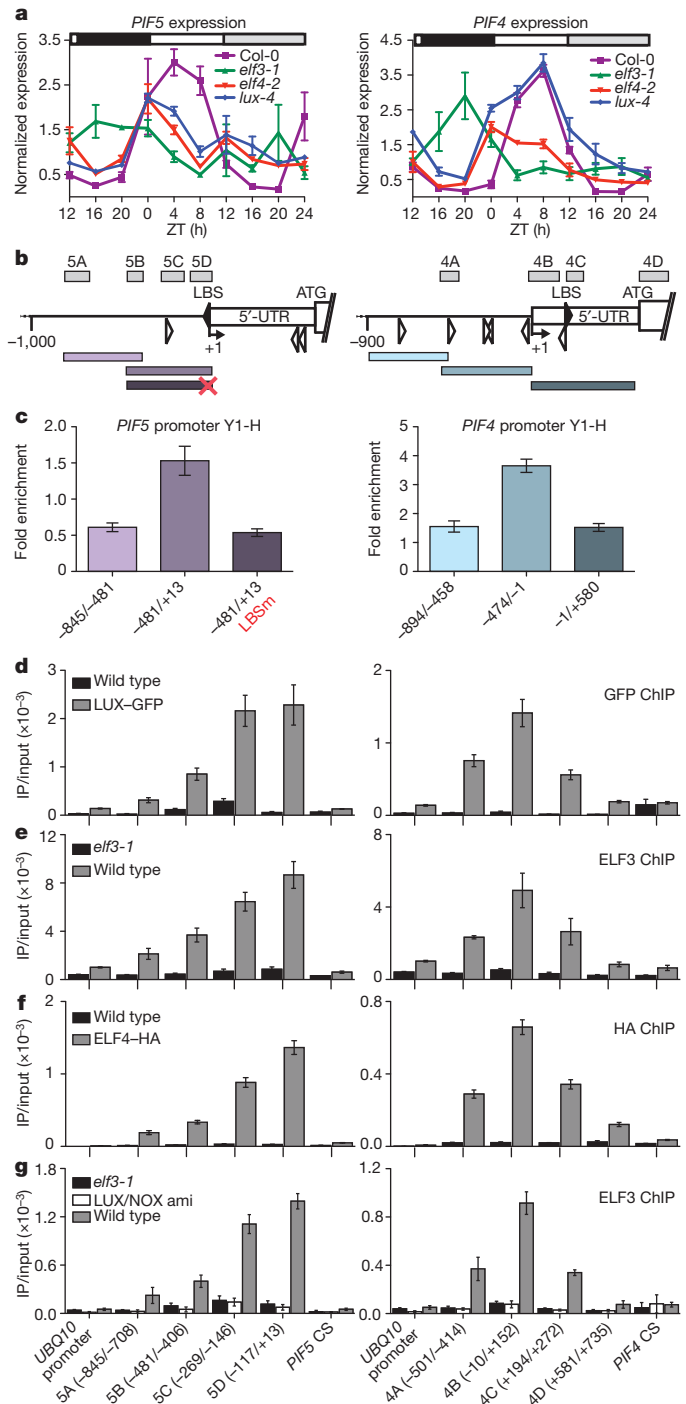


transcriptional repressors in the circadian clock<sup>20,24</sup>. The levels of *PIF4* and *PIF5* mRNA are elevated in *elf3-1*, *elf4-2* and *lux-4* mutants compared with the wild type, particularly during the early evening (Fig. 3a). Recent work demonstrated that the addition of an activation domain to LUX (LUX-VP64) induced a neomorphic hypocotyl elongation phenotype<sup>20</sup>, and we found that *PIF4* and *PIF5* expression levels were increased in this background (Supplementary Fig. 7). These results, as well as the presence of full consensus LUX-binding sites (LBSs)<sup>20</sup> in the 5'-untranslated region of both *PIF4* and *PIF5* (Fig. 3b), suggested that LUX may participate directly in the modulation of *PIF4* and *PIF5* expression. Indeed, LUX was able to directly bind to the *PIF4* and *PIF5* promoters in yeast, and LUX binding to the *PIF5* promoter (from -481 to +13 base pairs) was lost when the consensus LBS was mutated (Fig. 3c).

To determine whether components of the EC were bound to the *PIF4* and *PIF5* promoters *in vivo*, chromatin immunoprecipitation (ChIP) assays were performed in *LUX::LUX-GFP* transgenic lines and then the *PIF4* and *PIF5* promoter sequences were amplified. These experiments revealed *in vivo* binding to the LBS in the promoters of *PIF4* and *PIF5* but not to control sequences in the coding regions of these genes or in the *POLYUBIQUITIN10* (*UBQ10*) promoter (Fig. 3d). The formation of the EC (Fig. 2) suggested that all of its components might participate in the regulation of *PIF4* and *PIF5* expression; therefore, we performed similar ChIP experiments for *ELF3* and *ELF4-HA*. We found that *ELF3* and *ELF4-HA* showed specific enrichment at the *PIF4* and *PIF5* promoter sequences that were bound by LUX (Fig. 3e, f). Additionally, *ELF3* ChIP experiments performed at the trough of EC levels (ZT2) showed a lower specific enrichment than those performed at ZT14 (Supplementary Fig. 8).

The localization pattern of the EC components on the *PIF4* and *PIF5* promoters suggested that the transcription factor LUX might be responsible for recruitment. *ELF3* ChIP experiments in *lux-4* seedlings demonstrated that less *ELF3* was recruited to the *PIF4* and *PIF5* promoters in these mutants but that recruitment was not completely abrogated (Supplementary Fig. 9). Previous work identified a MYB-domain-containing transcription factor highly similar to LUX, named NOX (At5g59570)<sup>11,12,20</sup>. NOX binds sequences that are similar to those bound by LUX in yeast<sup>20</sup> and was also able to form a complex with *ELF4* and *ELF3* (Supplementary Fig. 10a). We designed an artificial microRNA (amiRNA) using a web-based amiRNA algorithm (<http://wmd3.weigelworld.org/cgi-bin/webapp.cgi>) and generated an amiRNA-transgenic line in which the levels of both NOX and LUX would simultaneously be reduced (denoted LUX/NOX ami)<sup>25,26</sup>. LUX protein and NOX expression levels were reduced in this line (Supplementary Fig. 10b, c), which showed similar defects in circadian rhythms to *lux-4* mutants (Supplementary Fig. 10e, f); however, we observed an increase in hypocotyl length and *PIF4* and *PIF5* expression level compared with *lux-4* mutants (Supplementary Fig. 10d, g). When *ELF3* ChIP assays were performed in the LUX/NOX ami line, we observed a loss of the *ELF3* signal at the *PIF4* and *PIF5* promoters (Fig. 3g). *ELF3* was still present in extracts from these plants (Supplementary Fig. 10c), suggesting that the recruitment of *ELF3* (and therefore the EC) is mediated by both LUX and NOX.

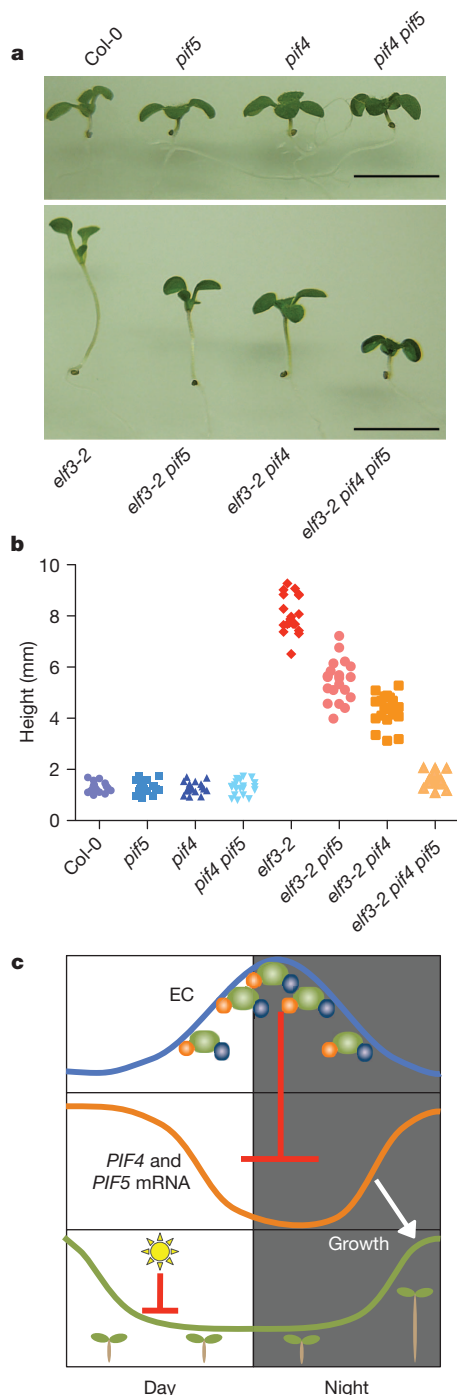
Previous reports showed that ectopic overexpression of the MYB-domain-containing transcription factors encoded by *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*; in the *CCA1-OX* line) or *LATE ELONGATED HYPOCOTYL* (*LHY*; in *lhy-1* mutants) resulted in phenotypes similar to those of *elf3*, *elf4* or *lux-4* (Fig. 3a and Supplementary Fig. 11). As *CCA1* and *LHY* form a complex that controls the expression of evening-element-containing genes<sup>27</sup>, such as *ELF4* and *LUX*<sup>11,28</sup>, the misexpression of *PIF4* and *PIF5* seen in *CCA1-OX* or *lhy-1* lines could be a result of EC misregulation. Therefore, we analysed the expression of *ELF4*, *ELF3* and *LUX* in the *lhy-1* background using the DIURNAL database<sup>18,19</sup>. We found that *ELF4* was clamped low, whereas *ELF3* and *LUX* were shifted 4 h and 12 h later, respectively (Supplementary Fig. 11). These results are consistent with the



**Figure 3 | The EC regulates *PIF5* and *PIF4* expression through recruitment by LUX.** **a**, *PIF5* and *PIF4* expression in *elf3*, *elf4*, *lux* and wild type (Col-0). Rectangles indicate light conditions as denoted in Fig. 1a. Error bars, s.e.m.;  $n = 3$ . **b**, *PIF5* (left) and *PIF4* (right) promoters denoting a degenerate (GATWCK or GATWYG) or consensus (GATWCG) LBS, represented by unfilled or filled arrowheads, respectively. Numbers are relative to the transcriptional start site (+1). Rectangles represent ChIP amplicons (top) and fragments for yeast one-hybrid assays (bottom); the red X denotes a mutated LBS. UTR, untranslated region. **c**, Yeast one-hybrid (Y1-H) with LUX-GAL4-AD and *PIF5* and *PIF4* promoter fragments (where / denotes a range). The fold enrichment is relative to controls. LBSm, LBS mutant. **d–g**, ChIP on *PIF5* and *PIF4* at ZT14, under extended light conditions: LUX (**d**), *ELF3* (**e**, **g**) and *ELF4* (**f**). **d**, **e**, Data are presented as mean  $\pm$  s.e.m.;  $n = 3$ . **f**, **g**, Data are presented as mean  $\pm$  s.d. (from two technical replicates measured twice). Experiments were repeated with similar results. CS, coding sequence.

circadian clock having a crucial role in the proper expression and phasing of the EC proteins.

If improper regulation of *PIF4* and *PIF5* underlies the hypocotyl growth defects observed in the EC mutants, then loss of *PIF4* and *PIF5*



**Figure 4 | Hypocotyl growth defects are rescued by loss of *PIF5* and *PIF4* in EC component mutant backgrounds.** **a**, Growth defects in the *elf3-2* background require *PIF4* and *PIF5*. Scale bar, 5 mm. **b**, Scatter plot of hypocotyl measurements from the wild type (Col-0), as well as *elf3-2*, *pif4-101* and *pif5-1* single and compound mutants. This experiment was repeated with similar results. **c**, The model represents the action of the EC on *PIF4* and *PIF5* expression during the early evening, which results in the gating of hypocotyl growth in *A. thaliana* seedlings. The circadian-regulated EC represses *PIF4* and *PIF5* expression in the evening. Throughout the day, post-transcriptional light-mediated degradation of *PIF4* and *PIF5* proteins inhibits growth. Near dawn, the concomitant rise in *PIF4* and *PIF5* mRNA and *PIF4* and *PIF5* protein levels promotes growth (white arrow).

should be epistatic to loss of the EC. To test this, we introduced *pif4* and *pif5* mutant alleles into the *elf3-2* mutant background, because mutating *ELF3* caused dissolution of the EC (Fig. 2a). Loss of *PIF4* or *PIF5* additionally mitigated the hypocotyl length defect in *elf3-2* (Fig. 4a, b), indicating that the hypocotyl phenotypes of EC mutants are mainly caused by misexpression of *PIF4* and *PIF5*. In addition, loss of *PIF4* and/or *PIF5* did not restore circadian rhythms in an *elf3* background (Supplementary Fig. 12), consistent with *PIF4* and *PIF5* being clock outputs that do not feed back into the oscillator<sup>4</sup>.

In summary, we have identified a novel multiprotein complex that directly links the circadian clock to diurnal regulation of hypocotyl growth. The ELF4–ELF3–LUX complex is regulated by the clock and by light (Figs 1a and 2b–d) and represses the expression of *PIF4* and *PIF5* in the early evening (Fig. 4c). This process is combined with light-regulated turnover of *PIF4* and *PIF5*, allowing maximum hypocotyl growth at dawn under diurnal conditions<sup>4,13</sup> (Fig. 4c). *ELF3* is necessary and sufficient to bring together *ELF4* and *LUX* to form a complex (Figs 1d and 2a), providing a mechanistic framework for understanding the shared phenotypes of EC component mutants in regulating circadian rhythms, growth and flowering. The role of *ELF3* as an adaptor protein is similar to its previously described capacity to modulate GIGANTEA levels through association with CONSTITUTIVELY PHOTOMORPHOGENIC 1 to regulate flowering and circadian rhythms<sup>22</sup>. The EC is composed of multiple proteins that are known to regulate signalling from the environment<sup>4,5,9–12,14–17,20,22,24,28</sup>; therefore, elucidating EC function will ultimately contribute to understanding how biochemical, physiological and developmental outputs are gated by the clock.

## METHODS SUMMARY

All wild-type, mutant and transgenic lines were in the *A. thaliana* ecotype Columbia-0 (Col-0). All transgenic and mutant lines were brought to homozygosity before use. The procedures for *A. thaliana* husbandry, yeast one-hybrid, two-hybrid and three-hybrid analyses, bioluminescent imaging, immunoprecipitation assays, ChIP assays and hypocotyl measurements have been described previously<sup>20,29,30</sup> and were carried out with modifications detailed in the Methods. In all growth chambers, light was supplied at  $80 \mu\text{mol m}^{-2} \text{s}^{-1}$  by cool-white fluorescent bulbs at 22 °C. For yeast two-hybrid analyses, SD-WL medium was used to select for the presence of both bait and prey vectors, and SD-WLHA medium was used to select for an interaction between the bait and the prey proteins. *IPP2*, *APX3* and *Atlg11910* levels were used to normalize real-time PCR expression analyses, and all primers for quantitative PCR are listed in Supplementary Table 1. The *ELF4::ELF4-HA* construct includes a 580-bp promoter sequence cloned from Col-0 DNA that was amplified using primers listed in Supplementary Table 1. The sequence TATGATATCCTTGCGTACCCA is the target of the LUX/NOX ami. Antibodies were generated in rabbits (Sigma Genosys) against either an ELF3-specific peptide (CSIQEERKRYDSSKP) or a full-length LUX protein fused to glutathione S-transferase (GST). Antibodies were affinity purified against the same ELF3-specific peptide using a SulfoLink Immobilization Kit (Thermo Scientific) or a GST–LUX affinity column. All immunoprecipitations were performed with Protein G Dynabeads (Invitrogen). For western blotting, ACTIN served as a loading control. Blots for ELF4 represent 20% of the total immunoprecipitation sample, because ELF4 needed to be separated on a different, 15%, gel, owing to its low molecular weight. Hypocotyl measurements were performed on evenly spaced seedlings grown under a 12L:12D cycle and measured on day 10.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 15 October 2010; accepted 6 May 2011.

Published online 13 July 2011.

- Wijnen, H. & Young, M. W. Interplay of circadian clocks and metabolic rhythms. *Annu. Rev. Genet.* **40**, 409–448 (2006).
- Yakir, E., Hilman, D., Harir, Y. & Green, R. M. Regulation of output from the plant circadian clock. *FEBS J.* **274**, 335–345 (2007).
- Harmer, S. L. The circadian system in higher plants. *Annu. Rev. Plant Biol.* **60**, 357–377 (2009).
- Nozue, K. *et al.* Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448**, 358–361 (2007).
- Michael, T. P. *et al.* A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol.* **6**, e225 (2008).

6. Niwa, Y., Yamashino, T. & Mizuno, T. The circadian clock regulates the photoperiodic response of hypocotyl elongation through a coincidence mechanism in *Arabidopsis thaliana*. *Plant Cell Physiol.* **50**, 838–854 (2009).
7. de Montaigu, A., Tóth, R. & Coupland, G. Plant development goes like clockwork. *Trends Genet.* **26**, 296–306 (2010).
8. Zagotta, M. T. *et al.* The *Arabidopsis* *ELF3* gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering. *Plant J.* **10**, 691–702 (1996).
9. Hicks, K. A., Albertson, T. M. & Wagner, D. R. *EARLY FLOWERING3* encodes a novel protein that regulates circadian clock function and flowering in *Arabidopsis*. *Plant Cell* **13**, 1281–1292 (2001).
10. Doyle, M. R. *et al.* The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. *Nature* **419**, 74–77 (2002).
11. Hazen, S. P. *et al.* *LUX ARRHYTHMO* encodes a Myb domain protein essential for circadian rhythms. *Proc. Natl Acad. Sci. USA* **102**, 10387–10392 (2005).
12. Onai, K. & Ishiura, M. *PHYTOCLOCK 1* encoding a novel GARP protein essential for the *Arabidopsis* circadian clock. *Genes Cells* **10**, 963–972 (2005).
13. Lorrain, S., Allen, T., Duek, P. D., Whitelam, G. C. & Fankhauser, C. Phytochrome-mediated inhibition of shade avoidance involves degradation of growth-promoting bHLH transcription factors. *Plant J.* **53**, 312–323 (2008).
14. Hicks, K. A. *et al.* Conditional circadian dysfunction of the *Arabidopsis* early-flowering 3 mutant. *Science* **274**, 790–792 (1996).
15. Liu, X. L., Covington, M. F., Fankhauser, C., Chory, J. & Wagner, D. R. *ELF3* encodes a circadian clock-regulated nuclear protein that functions in an *Arabidopsis* PHYB signal transduction pathway. *Plant Cell* **13**, 1293–1304 (2001).
16. Khanna, R., Kikis, E. A. & Quail, P. H. *EARLY FLOWERING 4* functions in phytochrome B-regulated seedling de-etiolation. *Plant Physiol.* **133**, 1530–1538 (2003).
17. Thines, B. & Harmon, F. G. Ambient temperature response establishes *ELF3* as a required component of the core *Arabidopsis* circadian clock. *Proc. Natl Acad. Sci. USA* **107**, 3257–3262 (2010).
18. Mockler, T. C. *et al.* The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 353–363 (2007).
19. Michael, T. P. *et al.* Network discovery pipeline elucidates conserved time-of-day-specific *cis*-regulatory modules. *PLoS Genet.* **4**, e14 (2008).
20. Helfer, A. *et al.* *LUX ARRHYTHMO* encodes a nighttime repressor of circadian gene expression in the *Arabidopsis* core clock. *Curr. Biol.* **21**, 126–133 (2011).
21. Hazen, S. P. *et al.* Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol.* **138**, 990–997 (2005).
22. Yu, J. W. *et al.* COP1 and *ELF3* control circadian function and photoperiodic flowering by regulating GI stability. *Mol. Cell* **32**, 617–630 (2008).
23. de Lucas, M. *et al.* A molecular framework for light and gibberellin control of cell elongation. *Nature* **451**, 480–484 (2008).
24. Dixon, L. E. *et al.* Temporal repression of core circadian genes is mediated through *EARLY FLOWERING 3* in *Arabidopsis*. *Curr. Biol.* **21**, 120–125 (2011).
25. Schwab, R., Ossowski, S., Rieger, M., Warthmann, N. & Weigel, D. Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* **18**, 1121–1133 (2006).
26. Ossowski, S., Schwab, R. & Weigel, D. Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* **53**, 674–690 (2008).
27. Lu, S. X., Knowles, S. M., Andronis, C., Ong, M. S. & Tobin, E. M. CIRCADIAN CLOCK ASSOCIATED1 and LATE ELONGATED HYPOCOTYL function synergistically in the circadian clock of *Arabidopsis*. *Plant Physiol.* **150**, 834–843 (2009).
28. Kikis, E. A., Khanna, R. & Quail, P. H. *ELF4* is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY. *Plant J.* **44**, 300–313 (2005).
29. Sawa, M., Nusinow, D. A., Kay, S. A. & Imaizumi, T. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* **318**, 261–265 (2007).
30. Para, A. *et al.* *PRR3* is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* **19**, 3462–3473 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** P. Y. Pongsawakul illustrated the model in Fig. 4c. We thank G. Breton, T. Hirota, J. Pruneda-Paz, D. Nagel, E. Kolmos and B. J. Cole for critical reading of the manuscript. We also thank J. Halverson, A. L. Quiroz and C. Valdivia for excellent technical assistance. *pif4-101 pif5-1* seedlings were a generous gift from S. Lorrain and C. Fankhauser. F. Harmon originally identified the nature of the *elf4-3* mutation and designed the dCAPS strategy. This work was supported by a University of California, San Diego, Chancellor's Undergraduate Research Scholarship (to J.J.K.), grants from the European Molecular Biology Organization (ALTF 236-2005 to A.H.), the National Science Foundation (IBN-0416762 to T.F.S.) and the National Institutes of Health (NRSA GM083585 to D.A.N., NRSA GM080930 to E.E.H., R01 GM79712 to T.I., and R01 GM50006 and GM67837 to S.A.K.).

**Author Contributions** D.A.N., A.H., E.E.H., T.I., T.F.S., E.M.F. and S.A.K. designed the experiments. D.A.N. performed and analysed all of the immunoprecipitations and ChIP assays, generated and characterized the anti-*ELF3* antibody and the transgenic lines, and generated the plasmids for yeast three-hybrid analysis. D.A.N. also co-performed the gene expression analysis with A.H. A.H. performed the yeast two- and one-hybrid assays and generated and characterized the LUX and LUX/NOX ami transgenic lines. E.E.H. generated the anti-LUX antibody, characterized transgenic lines and co-performed western blot analysis with D.A.N. J.J.K. measured hypocotyls, performed and analysed the yeast three-hybrid assay, and assisted with the yeast one-hybrid analysis and generation of the *elf3-2 pif4-101 pif5-1* seedlings. T.I. performed the original yeast two-hybrid assays. T.F.S. generated the *ELF4::ELF4-HA elf4-2* line. E.M.F. characterized the *ELF4*-transgenic lines. D.A.N. and S.A.K. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.A.K. ([skay@ucsd.edu](mailto:skay@ucsd.edu)).



## METHODS

**Yeast one-hybrid analysis.** All reporter strains were generated by homologous recombination of pGLacZi constructs (Clontech) in the yeast strain YM4271, according to the manufacturer's instructions. pGLacZi is a Gateway-compatible version of pLacZi (Clontech)<sup>20</sup>. Promoter fragments were amplified using primers listed in Supplementary Table 1 and were cloned into pENTR/D-TOPO (Invitrogen) and then transferred to pGLacZi, according to the manufacturer's instructions. To generate translational fusions to GAL4-AD, the coding sequence of *LUX* was cloned into pENTR/D-TOPO and subsequently recombined into pACTGW as previously described<sup>20</sup>. Transformations of AD constructs into the reporter strains and determinations of the  $\beta$ -galactosidase ( $\beta$ -gal) activity were performed in a 96-well format as previously described<sup>20</sup>.  $\beta$ -Gal activities were normalized to the control with an empty pACTGW vector.

**Yeast two-hybrid analysis.** cDNAs encoding full-length *LUX* (described above), *ELF3*, *ELF4*, *LUX-N* (amino acids 1–143) and *LUX-C* (amino acids 144–324) were cloned into the pENTR/D-TOPO vector (Invitrogen) (Supplementary Table 1). After the sequences had been verified, they were transferred into the pACTGW vector by Gateway LR recombination reaction (Invitrogen) to generate the bait plasmids<sup>20</sup>. *ELF4* and *ELF3* cDNAs were transferred into pASGW by a Gateway LR recombination reaction (Invitrogen) to generate the prey plasmids. The detailed yeast two-hybrid procedure was as previously described<sup>29</sup>.

**Yeast three-hybrid analysis.** Yeast three-hybrid analysis was performed as described previously<sup>30</sup>, with the following modifications: *ELF3* with an amino-terminal FLAG-epitope tag was cloned from cDNA into a pENTR/D-TOPO vector using the primers 5'-CGCGGCCGCAATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAAATGAAGAGAGGGAAAGATGAGGAG-3' and 5'-TTGGTTCTGCCATGAGACTG-3', and then inserted into the original pENTR/dTOPO-*ELF3* clone using the restriction enzymes NotI and EcoRI (New England BioLabs) and confirmed by sequencing. *ELF4* was then cloned into the pBridge vector by amplifying with the primers 5'-GGGGGAATTCATGAAGAGGAACGGCGAGAC-3' and 5'-TTTTCTGCAGTTAAGCTCTAGTTCCGGCAGC-3', and inserting into EcoRI and PstI (New England BioLabs) restriction sites. *FLAG-ELF3* was then cloned into either the pBridge vector or pBridge-*ELF4* using the restriction sites of NotI and EcoRV (New England BioLabs), after first digesting either the pBridge or pBridge-*ELF4* vector with BglI, blunting with Klenow and then digesting with NotI (New England BioLabs). pBridge-*ELF3* or pBridge-*ELF4-ELF3* was introduced into yeast strain YM4271 and then mated to strains containing the vector pACTGW, pACTGW-*LUX* or pACTGW-*NOX*<sup>20</sup> in the yeast strain AH109, according to the manufacturer's protocol (Clontech). Yeast were grown under selection and analysed for  $\beta$ -gal activity, as described by the manufacturer's instructions (Clontech) with the modifications for 96-well analysis<sup>20</sup>.

**Plant materials and growth conditions.** All wild-type, mutant and transgenic lines were in *A. thaliana* ecotype Columbia-0. *CAB2::LUC*-reporter-containing lines have been described previously<sup>31</sup>. Seeds were chlorine-gas sterilized and plated onto 1× Murashige and Skoog (MS) basal salt medium with 1.5% agar and 3% (w/v) sucrose. After stratification in the dark at 4 °C for 3 days, plates were transferred to an incubator (Percival Scientific) that was set to the indicated light conditions and a constant temperature of 22 °C. Light entrainment was in 12L:12D cycles or in short-day and long-day photoperiods (8L:16D and 16L:8D, respectively), with light supplied at 80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  by cool-white fluorescent bulbs. To analyse seedling morphology, evenly spaced seedlings were grown under 12L:12D conditions at 22 °C and measured on day 10. Photographs of seedlings were analysed using NIH ImageJ software (<http://rsbweb.nih.gov/ij/>).

**Construction of double and triple mutants.** *ELF4::ELF4-HA elf3-1 elf4-2 CAB2::LUC* double mutants were generated by genetic crosses between *elf3-1* (ref. 14) and *elf4-2 ELF4::ELF4-HA #1* (Basta resistance) *CAB2::LUC*, and *F<sub>2</sub>* populations were screened for long hypocotyls, Basta resistance, luminescence and an arrhythmic bioluminescence phenotype in constant light. *elf4-2 (arr44)*<sup>21</sup> mutations were identified by dCAPS PCR method<sup>32</sup> using the primers 5'-ATGGGTTTGCTCCACGATTA-3' and 5'-CAGGTTCCGGGAACCAAATTCT-3', and the restriction enzyme HpyCH4V (New England BioLabs) to analyse for the presence of the mutation. The *elf3-1* mutation was confirmed by 100% long hypocotyl, as well as by analysis using dCAPS primers 5'-TTTGCAGAGGATAAGCTGCGCT-3', 5'-TGTTGGCTGTTGCTGTTGCTGT-3' and the restriction enzyme HincII, and by loss of the *ELF3* signal in western blotting. *lux-4 elf3-1 CAB2::LUC* double mutants were generated by crossing *elf3-1* to *lux-4 CAB2::LUC*, and *F<sub>2</sub>* populations were screened for long hypocotyls, luminescence and an arrhythmic bioluminescence phenotype in constant light. Loss of *LUX* and *ELF3* was confirmed by assessing hypocotyl length, performing dCAPS PCR for *elf3-1* and *lux-4* (using the primers 5'-ATGGAGATGACGGTGGCGGT-3' and 5'-AACGAATCTCTTGTTAGCTGCGGAGT-3' and the restriction enzyme HinfI (New England BioLabs)), and carrying out

western blot analysis. *elf4-3 elf3-1 CAB2::LUC* double mutants were generated by crossing *elf4-3 CAB2::LUC*, which was generated by EMS mutagenesis as previously described<sup>11</sup>, with *elf3-1*, and these mutants were screened as above. The mutant *elf4-3* contains a single point mutation in the coding sequence of *ELF4* that results in a truncated protein (W26\*), which was identified by sequencing. The dCAPS primers 5'-GAGCAGGGAGAGGATCCAGCGATGTG-3' and 5'-CCGACGAGAACTAGTATTGA-3' and the restriction enzyme BstXI (New England BioLabs) were used to screen for the mutation. The presence of the *elf3-1* mutation was confirmed by dCAPS and western blotting. The *elf3-2* lines<sup>14</sup> were crossed to *TOC1::LUC* lines as described previously<sup>33</sup>, and we analysed *F<sub>2</sub>* populations for long hypocotyls and bioluminescence. The *elf3-2* mutation was mapped using the TAIL PCR method, which identified an inversion<sup>34</sup>. A PCR strategy over the inversion was used to distinguish wild-type lines from mutant lines using the following primers: 5'-TGAGTATTTGTTTCTTCTCGAGC-3' and 5'-CATATGGAGGAAGTAGCCATTAC-3' for wild type and 5'-TGGTTATTTATTTCTCCGCTCTTTC-3' and 5'-TTGTTCCATTAGCTGTTCAACCTA-3' for *elf3-2*. The combination mutants *elf3-2 pif4-101 pif5-1 TOC1::LUC*, *elf3-2 pif5-1 TOC1::LUC*, *elf3-2 pif4-101 TOC1::LUC*, *pif4-101 pif5-1 TOC1::LUC*, *pif5-1 TOC1::LUC*, and *pif4-101 TOC1::LUC* were generated from crosses between *elf3-2 TOC1::LUC* and *pif4-101 pif5-1* double mutants. *F<sub>2</sub>* plants were screened for bioluminescence and then analysed for mutant backgrounds by PCR as previously described<sup>13</sup>. Homozygous *F<sub>3</sub>* populations were identified by screening for mutations and transgenes. The generation and characterization of *LUX::LUX-GFP lux-4 CAB2::LUC* has been described previously<sup>20</sup>.

**GFP and LUX/NOX ami line generation.** The coding sequence of GFP was amplified by PCR from the pK7FWG2 vector<sup>35</sup> using the following primers 5'-CACCATGTGGTCTCATCCTCAATTTGAAAAAGCGCGGTTGGTCTC ATCCTCAATTTGAAAAAGGTGGTATGGTGAGCAAGGCGAGGAGCTG-3' and 5'-TCAAGCGTAATCTGGAACATCGTATGGGTACACATCCTTGTA CAGCTCGTCCATGCC-3', which introduce a StrepII epitope (SII) tag to the N terminus and an HA tag to the carboxy terminus. This fragment was then cloned into Gateway pENTR/D-TOPO. After sequencing, this construct was recombined with the pB7WG2 vector<sup>35</sup> to constitutively express SII-GFP-HA under the control of the 35S promoter. This construct was introduced into *CCA1::LUC* lines<sup>36</sup> by *Agrobacterium*-mediated transformation<sup>37</sup>. Transformants were selected based on Basta resistance and fluorescence and were screened for single insertion. Lines were brought to homozygosity before use.

The amiRNA (TATGATATCCTTGCGTACCCA) targeting *LUX* and *NOX* was constructed as described previously<sup>25</sup>. Primers designed using WMD3 Web microRNA Designer (<http://wmd3.weigelworld.org/cgi-bin/webapp.cgi>) were used to amplify the amiRNA precursor by overlapping PCR from the pRS300 template. The fragment containing the amiRNA foldback was cloned into pENTR/D-TOPO, sequenced and subsequently recombined using Gateway LR Clonase II (Invitrogen) into the pB2GW7 vector<sup>35</sup> for constitutive expression under control of the 35S promoter. This construct was transformed into a *CAB2::LUC* reporter background<sup>31</sup> using *Agrobacterium* infiltration<sup>37</sup>. Transformants were selected on Basta, and all experiments were performed in single-insertion, homozygous plants.

**Luciferase imaging.** After 6 days of entrainment, plants were sprayed with 5 mM luciferin (Biosynth) prepared in 0.01% (v/v) Triton X-100 (Sigma-Aldrich) and transferred to constant light (80  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) 1 day before imaging. The emitted luminescence was recorded every 2.5 h over 5 days, using a digital CCD camera (Hamamatsu Photonics). The images were processed using MetaMorph imaging software (Molecular Devices), and the data were analysed by fast Fourier transform–nonlinear least squares (FFT–NLLS)<sup>38</sup> using the interface provided by the Biological Rhythms Analysis Software System version 3.0 (BRASS) (<http://www.amillar.org>).

**Generation of anti-*ELF3* antibody.** Antibodies were generated in rabbits (Sigma Genosys) against an *ELF3*-specific peptide, containing an additional N-terminal cysteine for conjugation (CSIQEERKRYDSSKP), corresponding to amino acids 681–694 of *ELF3*. Antibodies were affinity purified against this peptide using a SulfoLink Immobilization Kit (Thermo Scientific). Eluted antibody-containing fractions were buffer exchanged into 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 50% glycerol and 0.02% NaN<sub>3</sub> by using an equilibrated PD-10 column (GE Healthcare) and then stored at –80 °C.

**Generation of anti-*LUX* antibody.** Full-length *LUX* protein was expressed as a glutathione S-transferase (GST) fusion, which was purified and used to immunize rabbits to obtain polyclonal antisera (Open Biosystems). Antibodies were purified using an affinity column made of purified GST-*LUX* bound to Affi-Gel 15 Activated Immunoaffinity Support (Bio-Rad)<sup>39</sup>. Antibodies were eluted from the affinity column with 100 mM glycine, pH 2.5, exchanged into storage buffer (1× PBS, 50% glycerol and 0.02% NaN<sub>3</sub>) using a PD-10 buffer exchange column and then stored at –80 °C.

**Construction of *ELF4::ELF4-HA*.** *ELF4* was cloned from genomic DNA to include 580 bp of promoter sequence, using the primers 5'-CACCGTCTTGC ATAACATGAAGC-3' and 5'-AGCTCTAGTTCGGCAGCACC-3', and then cloned into Gateway pENTR/D-TOPO. After sequencing, this construct was recombined with pEarleyGate 301 to introduce a C-terminal HA tag to *ELF4* (ref. 40). This construct was introduced into *elf4-2 CAB2::LUC* lines by *Agrobacterium*-mediated transformation<sup>37</sup>. Transformants were selected based on Basta resistance and screened for single insertion. Lines were brought to homozygosity before use.

**ChIPs.** Roughly 5 g (fresh weight) whole seedlings were harvested and crosslinked for 10 min under vacuum in crosslinking buffer (10 mM Tris, pH 8.0, 1 mM EDTA, 250 mM sucrose, 1 mM PMSF and 1% formaldehyde). Crosslinking was quenched in 125 mM glycine, pH 8.0, under vacuum for 5 min, and then seedlings were washed three times in double-distilled water and rapidly frozen before disruption in a ball mill (Retsch) under liquid nitrogen. Ground tissue was processed as described previously<sup>41</sup>, with the following modifications: sucrose-gradient-purified nuclei were resuspended in SII buffer (100 mM Na-phosphate, pH 8.0, 150 mM NaCl, 5 mM EDTA, 5 mM EGTA, 0.1% Triton X-100, 1 mM PMSF and 1× protease inhibitor cocktail (Roche)) and sonicated (Branson) at 15% power, with 0.5 s on/off cycles for a total of 30 s on ice until the average chromatin size was ~ 500 bp. The extracts were clarified by centrifugation at 20,000g and stored at -80 °C until use. Technical replicates containing approximately 1.5 mg DNA were resuspended in 800 µl SII buffer, incubated with 2 µg anti-GFP antibody (ab290, Abcam), anti-HA antibody (3F10, Roche) or anti-ELF3 antibody bound to Protein G Dynabeads (Invitrogen) for 1.5 h at 4 °C and then washed five times with SII buffer. Chromatin was eluted from the beads twice at 65 °C with Stop buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 20 mM EDTA and 1% SDS). RNase- and DNase-free glycogen (2 µg) (Boehringer Mannheim) was added to the input and eluted chromatin before they were incubated with DNase- and RNase-free proteinase K (Invitrogen) at 65 °C overnight and then treated with 2 µg RNase A (Qiagen) for 1 h at 37 °C. DNA was purified by phenol–chloroform extraction, followed by two serial ethanol precipitations. Quantitative PCR reactions of the technical replicates were performed using the CFX384 Real Time PCR Detection System (Bio-Rad), with the following PCR conditions: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 10 s at 55 °C and 20 s at 72 °C in a buffer consisting of 1× Ex Taq buffer (TaKaRa Bio), 0.5× SYBR Green (Molecular Probes), 5 nM fluorescein (Bio-Rad), 0.05% (v/v) Tween 20, 2.5% (v/v) DMSO, 25 µg ml<sup>-1</sup> BSA (New England BioLabs), 0.25 mM dNTPs, 250 nM primers and 1 U Taq DNA polymerase (BioPioneer). Primers used in this study are listed in Supplementary Table 1.

**Immunoprecipitations and western blots.** Approximately 500 mg whole seedlings were transferred to 2-ml tubes with three 3.2-mm stainless steel beads, and then frozen and disrupted in a ball mill under liquid nitrogen. After removing ~100 mg tissue for RNA analysis, ground tissue was resuspended in 400 µl SII buffer containing 1× phosphatase inhibitor cocktails 1 and 2 (Sigma) and 10 µM MG-132 (Peptides International) and sonicated twice at 10% power, with 0.5 s on/off cycles for a total of 20 s on ice. Extracts were then clarified by centrifugation at 4 °C, measured for protein concentration using 1× Bradford reagent (Bio-Rad) and normalized to 3 mg ml<sup>-1</sup> in SII buffer for western blots. For immunoprecipitations, extracts were diluted to 1.5 mg ml<sup>-1</sup> in SI buffer. Anti-HA antibody (4 µg) crosslinked to Protein G Dynabeads was added to extracts, rotated for 1.5 h at 4 °C, then washed 3× with SII buffer. Precipitated protein was eluted by heating beads at 65 °C for 5 min in 25 µl SDS–PAGE loading buffer. Protein levels were then analysed by western blot using either horse radish peroxidase (HRP)-conjugated 3F10 anti-HA (1:2,000, Roche), anti-ELF3 (1:750) or anti-LUX (1:750) antibody, followed by HRP-conjugated anti-rabbit secondary antibody (1:2,000, Pierce). The ACTIN loading control was detected using an anti-ACTIN mouse antibody, mAB1501 (1:2,000, Millipore), followed by alkaline-phosphatase-conjugated

anti-mouse secondary antibody (1:4,000, Promega). Blots for *ELF4* represent 20% of the total immunoprecipitation sample, because *ELF4* must be run on a separate, 15%, gel owing to its low molecular weight; these gels are noted by (\*). The dot (•) denotes a background signal arising from the crosslinked HA beads (data not shown). LUX runs as high- and low-molecular-weight isoforms, denoted by (–). When samples were collected in the dark, extracts were made and immunoprecipitations were assembled under a safe green light and protected from light until they were eluted in SDS–PAGE loading buffer before loading onto gels.

**Antibody crosslinking.** Antibody (2 µg) was crosslinked to 12 µl Protein G Dynabeads according to the manufacturer's instructions, with the following modifications: quenching of the dimethyl pimelimidate was performed with 0.2 M ethanolamine (pH 8.0), followed by two washes with 0.1 M glycine (pH 2.5) and neutralization with neutralization buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl and 0.01% Triton X-100), and the samples were then stored at -20 °C in storage buffer (50% glycerol, 50 mM Tris, pH 8.0, 150 mM NaCl, 0.01% Triton X-100 and 0.03% NaN<sub>3</sub>) until use.

**RNA extractions.** Seedlings were grown on Whatman filter paper atop MS plates under 12L:12D, 8L:16D or 16L:8D conditions and harvested on day 12, or were transferred to constant light on day 10 and harvested 2 to 3 days later. Total RNA was isolated using an RNeasy Plant Mini Kit (Qiagen). For cDNA synthesis, 1 µg total RNA was reverse-transcribed using the iScript cDNA synthesis kit (Bio-Rad). Synthesized cDNA was quantified by real-time quantitative PCR using the CFX-384 Real Time System (Bio-Rad), with the following PCR conditions: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C, 10 s at 55 °C and 20 s at 72 °C in a buffer consisting of 1× ExTaq buffer, 1× SYBR Green, 10 nM fluorescein (Bio-Rad), 0.1% (v/v) Tween 20, 5% (v/v) DMSO, 50 µg ml<sup>-1</sup> BSA, 0.25 mM dNTPs, 250 nM primers and 1 U Taq DNA polymerase. Isopentenyl pyrophosphate/dimethylallyl pyrophosphate isomerase (*IPP2*) (At3g02780), ascorbate peroxidase 3 (*APX3*) (At4g35000) and aspartyl protease family protein (At1g1910) were used as the normalization controls<sup>11,19</sup>. Primer sequences are shown in Supplementary Table 1 and were designed using Primer3 (ref. 42) or as described for *PIF4* and *PIF5* (ref. 43).

31. Millar, A. J., Carre, I. A., Strayer, C. A., Chua, N. H. & Kay, S. A. Circadian clock mutants in *Arabidopsis* identified by luciferase imaging. *Science* **267**, 1161–1163 (1995).
32. Neff, M. M., Turk, E. & Kalishman, M. Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* **18**, 613–615 (2002).
33. Alabadi, D. et al. Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* **293**, 880–883 (2001).
34. Liu, Y. G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463 (1995).
35. Karimi, M., Inzé, D. & Depicker, A. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195 (2002).
36. Pruneda-Paz, J., Breton, G., Para, A. & Kay, S. A. A functional genomics approach reveals CHE as a component of the *Arabidopsis* circadian clock. *Science* **323**, 1481–1485 (2009).
37. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
38. Plautz, J. D. et al. Quantitative analysis of *Drosophila* period gene transcription in living animals. *J. Biol. Rhythms* **12**, 204–217 (1997).
39. Harlow, E. & Lane, D. *Using Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratories Press, 1999).
40. Earley, K. W. et al. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
41. Haring, M. et al. Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**, 11 (2007).
42. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
43. Czechowski, T., Bari, R. P., Stitt, M., Scheible, W. R. & Udvardi, M. K. Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J.* **38**, 366–379 (2004).

# Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP

Petra Wollmann<sup>1,2</sup>, Sheng Cui<sup>1†</sup>, Ramya Viswanathan<sup>3</sup>, Otto Berninghausen<sup>1</sup>, Melissa N. Wells<sup>3</sup>, Manuela Moldt<sup>1</sup>, Gregor Witte<sup>1,2,4</sup>, Agata Butryn<sup>1</sup>, Petra Wendler<sup>1</sup>, Roland Beckmann<sup>1,4</sup>, David T. Auble<sup>3</sup> & Karl-Peter Hopfner<sup>1,2,4</sup>

Swi2/Snf2-type ATPases regulate genome-associated processes such as transcription, replication and repair by catalysing the disruption, assembly or remodelling of nucleosomes or other protein–DNA complexes<sup>1,2</sup>. It has been suggested that ATP-driven motor activity along DNA disrupts target protein–DNA interactions in the remodelling reaction<sup>3–5</sup>. However, the complex and highly specific remodelling reactions are poorly understood, mostly because of a lack of high-resolution structural information about how remodellers bind to their substrate proteins. Mot1 (modifier of transcription 1 in *Saccharomyces cerevisiae*, denoted BTA1 in humans) is a Swi2/Snf2 enzyme that specifically displaces the TATA box binding protein (TBP) from the promoter DNA and regulates transcription globally by generating a highly dynamic TBP pool in the cell<sup>6,7</sup>. As a Swi2/Snf2 enzyme that functions as a single polypeptide and interacts with a relatively simple substrate, Mot1 offers an ideal system from which to gain a better understanding of this important enzyme family. To reveal how Mot1 specifically disrupts TBP–DNA complexes, we combined crystal and electron microscopy structures of Mot1–TBP from *Encephalitozoon cuniculi* with biochemical studies. Here we show that Mot1 wraps around TBP and seems to act like a bottle opener: a spring-like array of 16 HEAT (huntingtin, elongation factor 3, protein phosphatase 2A and lipid kinase TOR) repeats grips the DNA-distal side of TBP via loop insertions, and the Swi2/Snf2 domain binds to upstream DNA, positioned to weaken the TBP–DNA interaction by DNA translocation. A ‘latch’ subsequently blocks the DNA-binding groove of TBP, acting as a chaperone to prevent DNA re-association and ensure efficient promoter clearance. This work shows how a remodelling enzyme can combine both motor and chaperone activities to achieve functional specificity using a conserved Swi2/Snf2 translocase.

Mot1 is highly conserved among eukaryotes and consists of an amino-terminal TBP binding region of approximately 90–140 kDa with predicted HEAT repeats, followed by a carboxy-terminal Swi2/Snf2-type ATPase domain of approximately 60–70 kDa (refs 8, 9). To provide a structural framework for a remodeller–substrate complex, we determined the crystal structure of the N-terminal domain (NTD) of *Encephalitozoon cuniculi* (Ec) Mot1 (comprising the HEAT domain, residues 1–779, but lacking the ATPase domain, residues 780–1256) in complex with full-length EcTBP, to 3.1 Å resolution (Fig. 1 and Supplementary Table 1). EcMot1 has the characteristic sequence and biochemical features of *S. cerevisiae* (Sc) Mot1 and human BTA1, including TBP- and DNA-stimulated ATPase activity, TBP binding via its HEAT domain, and, most importantly, ATP-stimulated TBP displacement from TATA DNA (Supplementary Figs 1 and 2).

The EcMot1 NTD consists of a highly elongated stretch of 16 HEAT repeats, arranged in a horseshoe shape with dimensions of about 95 Å × 85 Å × 40 Å, and it forms a specific 1:1 complex with EcTBP (Fig. 1). Notably, Mot1 wraps around one side of the pseudosymmetric

TBP and grips both the convex protein-interacting surface and the concave DNA-binding surface of TBP via several loop insertions in the array of HEAT repeats. This wrapping interaction enables Mot1 to split the very stable EcTBP dimer that forms in the absence of DNA<sup>10</sup>, and that we observed biochemically and in a separate crystal structure of EcTBP alone at 1.9 Å resolution (Supplementary Fig. 3a and Supplementary Table 1). Despite this dual-sided grip, Mot1 does not alter the structure of TBP substantially because EcTBP bound to Mot1, EcTBP in the TBP dimer and ScTBP bound to DNA are all very similar (Supplementary Fig. 3c–f). This indicates that remodelling of TBP does not proceed via changes in TBP structure as a simple consequence of Mot1 binding, but requires the ATP-dependent action of the Swi2/Snf2 domain.

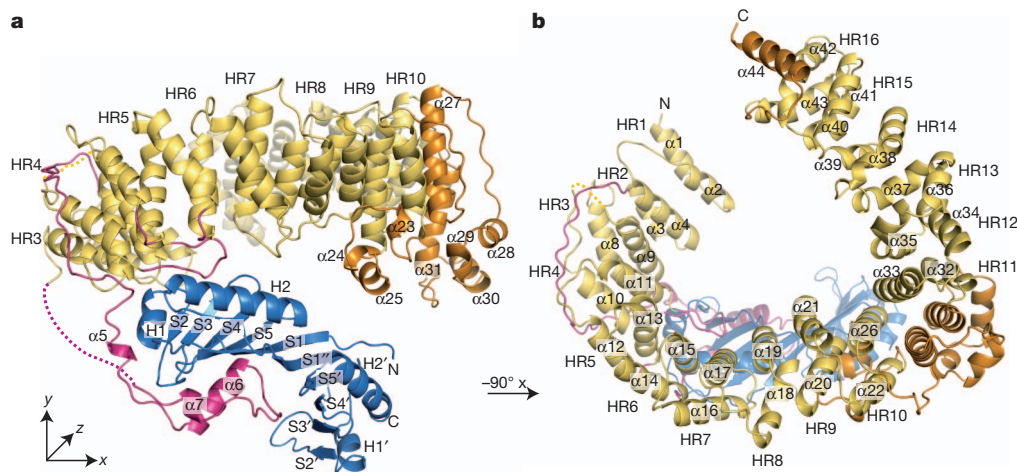
Promoter-bound TBP has its DNA-binding surface occupied, so Mot1 uses highly complementary HEAT-repeat loops to recognize the convex protein-interaction surface of TBP (Fig. 2a). In TBP, α-helices H1 and H2 are bound by the loop of HEAT repeat 4 (residues 209–221), and by interactions with α-helix 13 in HEAT repeat 5 and α-helix 15 in HEAT repeat 6. Most of these interactions are ion pairs between TBP R46 and Mot1 Q256, TBP R48 and Mot1 D212, TBP R65 and Mot1 D215, TBP R96 and Mot1 D216, TBP K99 and Mot1 D216, TBP K103 and Mot1 D290 and TBP K103 and Mot1 D292 (Supplementary Table 2). In addition, Mot1 F213 binds to a hydrophobic cleft between H1, H2 and β-sheet S2, providing a hydrophobic anchor, and Mot1 residues F210 and W255 pack against the side chains of TBP residues R48 and K103.

These interactions are well conserved evolutionarily (Supplementary Fig. 4a and Supplementary Table 2) and are supported by functional data *in vivo* and *in vitro*. For instance, ScTBP K145 (EcTBP K103) is an essential residue for stabilization of the ScMot1–ScTBP interaction<sup>8</sup>. We mutated K103 in EcTBP and observed that EcTBP(K103E) failed to form a stable complex with EcMot1(NTD) *in vitro* (Fig. 2b). Moreover, mutation of D365 (D212 in EcMot1) inactivated ScMot1 *in vivo* and abolished the Mot1–TBP interaction *in vitro*<sup>8</sup>. Mutations of K138 in ScTBP also impaired the interaction with ScMot1, consistent with the projection of the homologous side chains into the EcMot1(NTD)–EcTBP interface<sup>8,11</sup>. The distribution of residues along the length of the EcMot1 NTD is also consistent with earlier work showing that broad segments of the ScMot1 and BTA1 N termini are important for stable interaction with TBP<sup>8,9,12</sup>. Thus, the specific interaction interface between the Mot1 HEAT repeats and the convex surface of TBP is well suited to provide specific recognition of the TBP surface in the TBP–promoter complex, explaining why Mot1 specifically targets TBP–DNA and not other protein–DNA complexes.

Unexpectedly, the concave DNA-binding surface of TBP, accessible only when TBP is displaced from promoter DNA, is bound by Mot1 as well (Fig. 2c). A long ‘latch’, located between HEAT repeats 2 and 3, protrudes from the side of Mot1 distal to TBP and wraps all the way

<sup>1</sup>Department of Biochemistry, Ludwig-Maximilians-University, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. <sup>2</sup>Munich Center for Advanced Photonics, Gene Center Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. <sup>3</sup>Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, Virginia 22908, USA. <sup>4</sup>Center for Integrated Protein Science, Gene Center Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany. †Present address: Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 9, Dong Dan San Tiao, Beijing 100730, China.





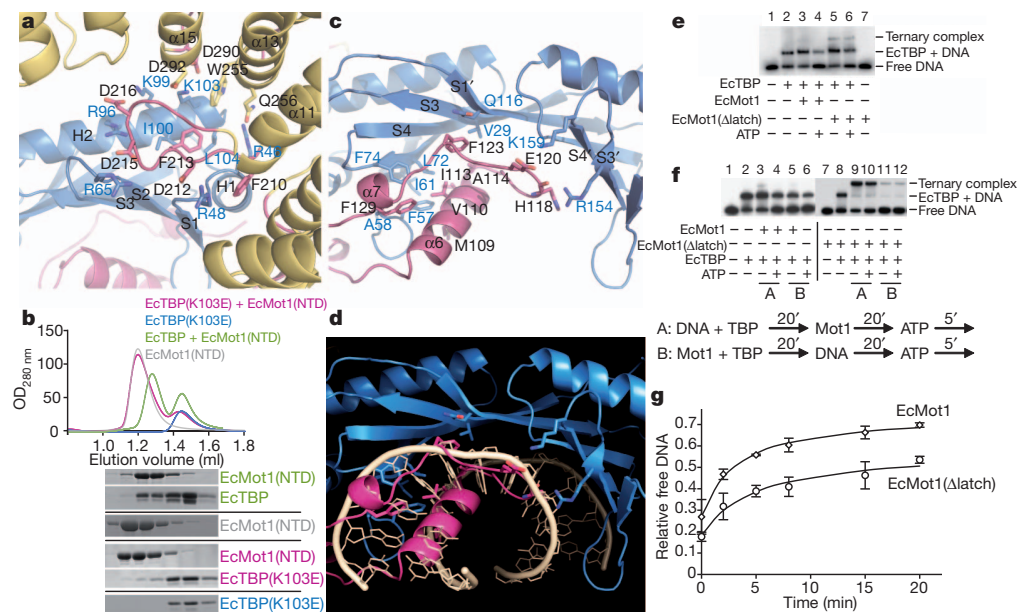
**Figure 1 | Overview of the EcMot1(NTD)-EcTBP structure.** **a, b,** Structure of the EcMot1(NTD)-EcTBP complex in ribbon representation with highlighted and annotated secondary structure. The HEAT repeats (HR) of EcMot1(NTD) are coloured yellow and non-HEAT-repeat insertions are in

orange. The latch and the loops of HR 4 to HR 6 are highlighted in magenta. EcTBP is coloured blue. Two loops not traced by electron density are indicated by dashed lines.

around the side of Mot1 and TBP. Notably, its tip (residues 101–130) substitutes for interactions made by four base pairs (bp) at and immediately downstream from the TATA sequence (Fig. 2d). A set of hydrophobic interactions matches the hydrophobic nature of TBP's DNA-binding groove. For instance, the side chain of M109 in Mot1 replaces a deoxyribose moiety in binding to TBP F57, a prominent and highly conserved DNA-binding residue of TBP. The main chain of residues 118–129 folds along the position of the backbone of the coding DNA strand, with side chains often placed at positions occupied by base and sugar moieties. F123 in Mot1 replaces a deoxyribose moiety

and stacks with the conserved TBP Q116, and F129 in Mot1 replaces a base moiety that interacts with the aromatic pair F57 and F74 in TBP.

To test the function of the latch, we generated the mutants EcMot1( $\Delta$ latch) and EcMot1(NTD $\Delta$ latch) that lack residues 96–132. Both proteins can still interact with EcTBP with approximately equal Mot1:TBP molarity (Supplementary Fig. 4b). This observation indicates that EcTBP is mainly bound by acidic loops of HEAT repeats 4–6 in Mot1. However, the latch might prevent TBP rebinding to DNA (after DNA dissociation) and might also prevent homodimerization by saturating the exposed, hydrophobic DNA-binding cleft of TBP (see



**Figure 2 | Details of the interaction interfaces and latch function.** **a,** Close-up view of the EcMot1-EcTBP interaction (colour scheme as in Fig. 1). **b,** Wild-type EcTBP and EcMot1(NTD) (green) can form a stable complex, whereas the EcTBP(K103E) mutant does not co-elute with EcMot1(NTD) (pink) in size exclusion chromatography (Supplementary Fig. 1b). **c, d,** The latch of EcMot1 (pink, shown in **c**) overlaps with the DNA-binding region (shown in **d**) of EcTBP (blue). Some bases of the superimposed DNA (wheat, from PDB 1YTB<sup>29</sup>) were omitted. **e, f,** Electrophoretic mobility shift assays (for corresponding quantifications see Supplementary Fig. 4). **e,** EcMot1( $\Delta$ latch) formed stable ternary complexes with EcTBP-DNA (lane 5). However,

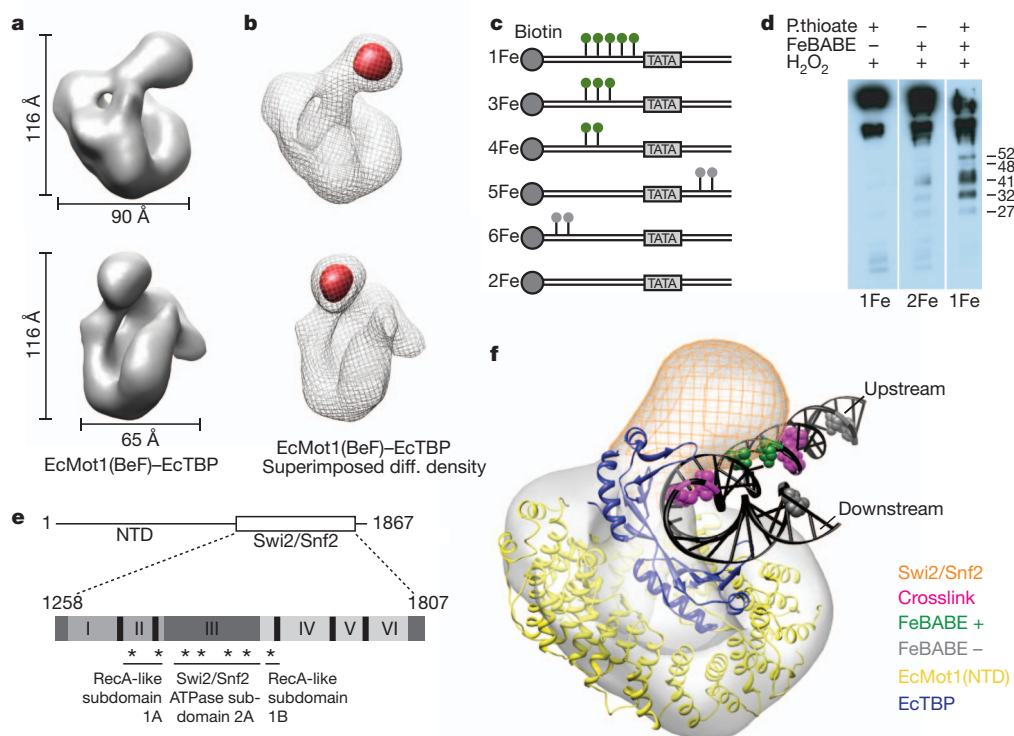
although wild-type EcMot1 largely cleared the DNA probe of bound TBP in an ATP-dependent reaction (lane 4), EcMot1( $\Delta$ latch) was less efficient in TBP removal (lane 6). **f,** EcMot1 was incubated with EcTBP after (A) or before (B) the addition of DNA. Preincubation of the two proteins inhibited TBP's ability to bind DNA. **g,** EcMot1( $\Delta$ latch) dissociated EcTBP-DNA less efficiently than wild-type EcMot1. ATP was added to pre-formed EcMot1-EcTBP-DNA or EcMot1( $\Delta$ latch)-EcTBP-DNA ternary complexes, and the proportion of free DNA was quantified by electrophoretic mobility shift assays at various times thereafter. Data represent mean and standard error from two independent experiments.

Supplementary Fig. 3b). Indeed, whereas EcMot1(NTD) forms a heterodimer with EcTBP, we found that EcMot1(NTD $\Delta$ latch) forms a 2:2 complex with EcTBP (Supplementary Table 3). The most likely explanation is that two EcMot1(NTD $\Delta$ latch) molecules bind the EcTBP dimer, but fail to dissociate the dimer owing to the absence of the latch. Because EcMot1( $\Delta$ latch) in complex with EcTBP does not show a substantially increased hydrodynamic radius compared to the wild-type complex in gel filtration (Supplementary Fig. 4b), it is likely that the Swi2/Snf2 domain sterically prevents dimerization of EcMot1( $\Delta$ latch) via TBP dimers.

Thus, although one function of the latch might be to keep TBP in a monomeric state, a more intriguing role might be to interfere with DNA binding by TBP. To test this, we analysed the ability of the EcMot1( $\Delta$ latch) protein to bind to the TBP–DNA complex. In contrast to wild-type EcMot1, EcMot1( $\Delta$ latch) formed readily detectable ternary complexes with EcTBP and DNA (Fig. 2e, f), indicating that the latch makes the association of EcMot1 with EcTBP–DNA less stable. Although it bound to TBP–DNA more efficiently, EcMot1( $\Delta$ latch) was notably impaired in ATP-dependent TBP–DNA dissociation (Fig. 2e–g and Supplementary Fig. 4d, e). This was not due to a defect in ATPase activity (Supplementary Fig. 4g). Moreover, when combined with EcTBP before DNA addition, EcMot1 inhibited DNA binding by EcTBP (Fig. 2f and Supplementary Fig. 4e). EcMot1(NTD) also inhibited DNA binding by EcTBP in a reaction that required the latch (Supplementary Fig. 4c, f). However, the latch was not essential for inhibiting the EcTBP–DNA interaction in the context of the full-length EcMot1 protein (Fig. 2f and Supplementary Fig. 4e), indicating that

both the latch and the ATPase domain can modulate EcTBP DNA-binding activity. Taken together, the data indicate that the latch has ‘chaperone’ activity and regulates macromolecular interactions with the hydrophobic groove of TBP. Because DNA binding and latch binding to TBP are mutually exclusive (Fig. 2d), it is unlikely that the latch initially disrupts the TBP–DNA complex. Consistent with this, EcMot1( $\Delta$ latch) was able to displace TBP from DNA using ATP, but the overall level of displacement was increased by the latch (Fig. 2g). Thus, our combined data can be explained by a physiologically plausible model in which the ATP-dependent action of the Swi2/Snf2 domain remodels TBP–TATA first, and then the latch blocks the exposed hydrophobic groove to prevent rebinding.

To reveal the architecture of the whole *E. cuniculi* Mot1–TBP complex, including its Swi2/Snf2 domain, we generated three-dimensional reconstructions of negatively stained EcMot1–EcTBP particles visualized in electron micrographs (Fig. 3a and Supplementary Fig. 5). The three-dimensional reconstruction is shaped like a slightly closed ‘C’ with a globular protrusion, and is similar to the three-dimensional reconstructions of the human TBP–BTAF1 complex<sup>13</sup>. To locate the Swi2/Snf2 domain unambiguously, we imaged a complex of EcTBP with a deletion mutant of EcMot1 in which the C-terminal half of the Swi2/Snf2 domain was truncated (EcMot1( $\Delta$ CT)) (Supplementary Figs 5 and 6c). We found that the prominent protrusion is missing from this complex, indicating that this protrusion corresponds to the C-terminal half of the ATPase (Fig. 3b). Finally, we imaged Mot1 without TBP (Supplementary Figs 5 and 6b). Although Mot1 alone is evidently more flexible than it is in the Mot1–TBP complex, and



**Figure 3 | Three-dimensional reconstruction of the EcMot1–EcTBP complex and model of the EcMot1–EcTBP–DNA complex.** **a**, Two views of the EcMot1(BeF)–EcTBP density map. ADP–BeF<sub>3</sub><sup>–</sup> was added owing to its assumed stabilization of the ATPase domain. **b**, Subtraction map (red) between EcMot1(BeF)–EcTBP (grey mesh) and EcMot1( $\Delta$ CT)–EcTBP density maps. **c**, Schematic of the DNA probes with phosphorothioates (green/grey lollipops) used in FeBABE cleavage assays. **d**, FeBABE-mediated cleavage of Mot1, analysed by western blot<sup>30</sup> with approximate sizes of the cleavage products in kilodaltons. **e**, Summary of FeBABE results. Asterisks represent approximate sites of cleavage mediated by FeBABE conjugated to the DNA upstream of the

TATA box. **f**, Model of the Mot1–TBP–DNA complex. Electron density map of EcMot1(BeF)–EcTBP complexes with the crystal structure of EcMot1(NTD)–EcTBP, including a superimposed elongated DNA from the ScTBP–DNA complex (PDB code 1YTB). Bases that represent 5-IdU substitutions used for crosslinking ScMot1 to DNA<sup>15</sup>, and bases that represent the FeBABE probe 4Fe (Supplementary Fig. 7a), are coloured in magenta and green, respectively. Positions of FeBABE conjugation that did not produce cleavage are coloured in grey. The position of the Swi2/Snf2 domain of Mot1 is indicated as an orange mesh.

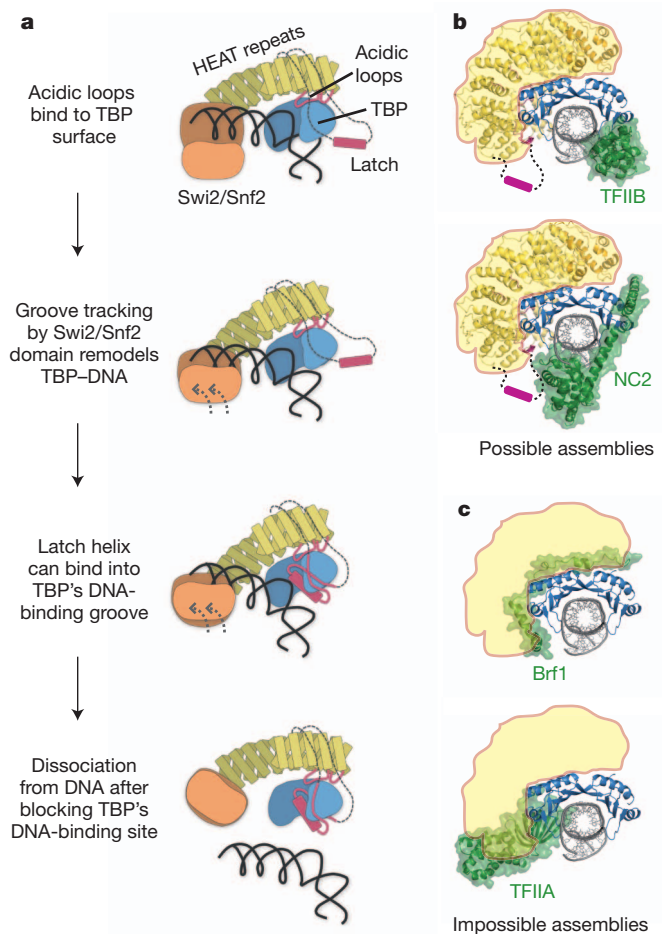


adopts a slightly different conformation, not unexpected for a large HEAT array, a particular lateral density patch was seen to be missing, thereby defining the location of TBP in the complex. Altogether, these data allowed us to rigid-body-dock the Mot1(NTD)-TBP crystal structure convincingly into the electron microscopy density (Supplementary Fig. 6a).

To corroborate this placement, we superimposed TBP in the crystal structure with the ScTBP-DNA complex, and extended the ends of the DNA with generic B-form DNA. Indeed, the upstream DNA protrudes towards the electron density corresponding to the Swi2/Snf2 domain in the electron microscopy three-dimensional reconstruction (Fig. 3f). Our model predicts that the Swi2/Snf2 domain contacts the DNA about 10–17 bases upstream from the TATA sequence, well positioned to translocate along the minor groove of the DNA<sup>3</sup>. This is in good agreement with previous crosslinking results and satisfactorily explains why a duplex DNA extension is required upstream of the TBP binding site for formation of a catalytically active ScMot1-TBP-DNA complex<sup>14,15</sup>. To validate this model further, we localized the region of ScMot1 proximal to the upstream DNA using FeBABE-mediated hydroxyl radical cleavage<sup>16</sup> (Fig. 3c and Supplementary Fig. 7a). As predicted by the model, FeBABE molecules positioned within a 9-bp DNA segment immediately upstream of the TATA sequence generated several specific C-terminal Mot1 fragments (cleavage in the Swi2/Snf2 domain), whereas no cleavage products were detected without FeBABE or when FeBABE molecules were conjugated to DNA upstream of this region or downstream of the TATA sequence (Fig. 3d, e and Supplementary Fig. 7b).

Our combined data indicate that Mot1 recognizes TATA-bound TBP by binding to the positively charged TBP surface at H1 and H2, and by binding of the Swi2/Snf2 domain to the minor groove of upstream DNA. We suggest that ATP-dependent groove tracking of the Swi2/Snf2 domain initially disrupts TBP-TATA, followed by binding of the latch to the exposed hydrophobic groove of TBP and full dissociation of Mot1-TBP from DNA (Fig. 4a). In this model, which is consistent with the translocation direction inferred for nucleosome remodelling enzymes<sup>17</sup>, the Swi2/Snf2 domain 'pulls' on TBP. Alternatively, the Swi2/Snf2 domain might push TBP. The precise tracking direction must await future studies, although the proposed two-step displacement could occur by translocation in either direction. In any case, the rotational force generated by tracking even a few base pairs of DNA by the Swi2/Snf2 domain could lift TBP from DNA sufficiently for the latch to bind. The energy of a few ATP-dependent translocation steps could be stored elastically in the HEAT repeats. In this way, Mot1 would act like a bottle opener to lift TBP from DNA, with the acidic loops functioning as the head, the HEAT repeats as the handle and the Swi2/Snf2 domain as the twisting hand.

Because TBP exists in many different complexes that could be substrates for the remodelling activity of Mot1, we compared the Mot1-TBP complex with other structurally characterized TBP complexes. The HEAT domain of Mot1 would be able to interact with TBP-TFIIB-DNA complexes as well as with TBP-NC2-DNA complexes (Fig. 4b). The compatibility of Mot1 and NC2 binding to TBP-DNA is consistent with several *in vitro* and *in vivo* results<sup>8,18–21</sup>, including recent genome-wide chromatin co-localization of Mot1 and NC2 (ref. 20). In contrast, Mot1 sterically overlaps with TFIIA, explaining how Mot1 and TFIIA compete for binding to TBP (Fig. 4c)<sup>10,11,22</sup>. Mot1 evidently also clashes with Brf1, a subunit of the Pol III initiation factor TFIIB (Fig. 4c), whereas we do not see any clashes with a recent TBP-TFIIB-Pol II preinitiation complex model (Supplementary Fig. 8)<sup>23,24</sup>. Thus, these comparisons indicate that Mot1 can act on specific subsets of preinitiation complexes in addition to TBP alone. These may include minimal and incomplete preinitiation complexes as well as NC2-repressed TBP complexes, whereas preinitiation complexes that include TFIIA and TBP-associated factors, or Pol III preinitiation complexes (containing Brf1), may be excluded from regulation by Mot1.



**Figure 4 | Proposed remodelling mechanism.** **a**, Proposed mechanism of Mot1-mediated displacement of TBP from the DNA. **b**, **c**, Models of possible Mot1 substrates, generated by superimposing the EcMot1-EcTBP crystal structure on other TBP-containing structures. **b**, Possible Mot1 (yellow) substrates are TBP complexes with TFIIB (PDB code 1AIS) and NC2 (PDB code 1JFI). The Mot1 latch is omitted from the structure, but is drawn as a cartoon. **c**, Sterically impossible Mot1 substrates are TBP complexes with TFIIA (PDB code 1NH2) or the TFIIB subunit Brf1 (PDB code 1NGM).

The discovery of the latch and its role in reducing DNA binding and TBP dimerization indicates that Mot1 not only displaces TBP, but blocks its hydrophobic surface patch to prevent interactions with DNA or other factors that bind to the concave surface. Mot1 thus acts as a TBP chaperone to control its interaction with other macromolecules. Mot1 might hold TBP in a diffusible state, explaining how it helps to redistribute TBP rapidly between different promoters and binding sites in the genome. Redistribution between promoters requires large diffusion steps between chromosomes and chromosome loops *in trans*, as opposed to sliding along DNA *in cis*, which is probably part of the repression mode of NC2 (ref. 25). This model is supported by the important role of Mot1 in the high cellular mobility of TBP<sup>6,7</sup> and by early findings that a substantial proportion of TBP resides in a stable complex with Mot1 in HeLa and yeast cell extracts<sup>26,27</sup>.

The unusual interactions between Mot1 and TBP might be necessary because of the high-affinity, hydrophobic DNA-binding mode of TBP, as well as the necessity for tight regulation of its binding to specific sites in the genome, while preventing nonspecific DNA interactions. Thus, a combination of motor and chaperone functions could be a more general feature of remodelling systems that deal with the assembly or disassembly of complexes between sticky proteins and DNA. In other systems, remodelling and chaperone functions may be provided by separate factors, as seen, for example, in the cooperation of the SWI/SNF nucleosome remodelling complex and the Asf1 histone chaperone<sup>28</sup>.



These results provide a high-resolution view of how a Swi2/Snf2-type remodeller interacts with its substrate; they show how the conserved ATP-dependent DNA translocase module can be used to generate high functional specificities within the large and diverse family of Swi2/Snf2 enzymes; and they provide a testable mechanism for a remodelling reaction.

## METHODS SUMMARY

Recombinant full-length EcMot1 (residues 1–1275), EcTBP, EcMot1( $\Delta$ CT) (residues 1–1016), EcMot1(NTD) (residues 1–779) and EcMot1( $\Delta$ latch) ( $\Delta$ 96–132) were produced in *E. coli* or insect cells. Protein purification was conducted using standard methods and proteins were crystallized by hanging-drop vapour diffusion. EcTBP crystals diffracted to 1.9 Å resolution and were measured at the Swiss Light Source (SLS). Native data of crystals from EcMot1(NTD)–EcTBP showed that they diffracted X-rays to 3.1 Å; these data were collected at the European Synchrotron Radiation Facility (ESRF). Data from derivative crystals of selenomethionine-labelled EcMot1(NTD)–EcTBP were collected to 3.3 Å at the SLS. The structure of EcTBP was solved by molecular replacement using yeast TBP (Protein Data Bank code 1TBP) as a search model. The structure of EcMot1(NTD)–EcTBP was determined using selenium single-wavelength anomalous dispersion in combination with molecular replacement, with the EcTBP structure as a partial model. EcMot1–EcTBP in the presence of 2 mM ADP and beryllium fluoride ( $\text{ADP-BeF}_3^-$ ), EcMot1(E912Q) (the Walker B mutant of EcMot1 was used instead of wild type owing to its enhanced stability) or EcMot1( $\Delta$ CT)–EcTBP were used for negative stain (2% uranyl acetate) electron microscopic studies. Micrographs were recorded on a Tecnai G2 Spirit TEM at 120 kV. Size-exclusion experiments were performed on Ettan LC system (GE Healthcare, Superose 12 PC 3.2/30). FeBABE (Dojindo) was conjugated to 68-bp DNA duplexes, based on the sequence of the adenovirus major late promoter. Biotinylation of the top strand's 5' end allowed the duplexes to be bound by streptavidin beads. After FeBABE conjugation, TBP and Mot1 were loaded onto the modified DNAs and cutting was initiated by addition of ascorbic acid and hydrogen peroxide.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 28 January; accepted 18 May 2011.**

**Published online 6 July 2011.**

- Cairns, B. R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193–198 (2009).
- Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
- Dürr, H., Korner, C., Müller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
- Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.* **16**, 2120–2134 (2002).
- Racki, L. R. *et al.* The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. *Nature* **462**, 1016–1021 (2009).
- Auble, D. T. The dynamic personality of TATA-binding protein. *Trends Biochem. Sci.* **34**, 49–52 (2009).
- de Graaf, P. *et al.* Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. *J. Cell Sci.* **123**, 2663–2671 (2010).
- Darst, R. P. *et al.* Mot1 regulates the DNA binding activity of free TATA-binding protein in an ATP-dependent manner. *J. Biol. Chem.* **278**, 13216–13226 (2003).
- Pereira, L. A., van der Knaap, J. A., van den Boom, V., van den Heuvel, F. A. & Timmers, H. T. TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. *Mol. Cell Biol.* **21**, 7523–7534 (2001).
- Pugh, B. F. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255**, 1–14 (2000).
- Auble, D. T. & Hahn, S. An ATP-dependent inhibitor of TBP binding to DNA. *Genes Dev.* **7**, 844–856 (1993).
- Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* **22**, 2994–3006 (2008).
- Pereira, L. A. *et al.* Molecular architecture of the basal transcription factor B-TFIID. *J. Biol. Chem.* **279**, 21802–21807 (2004).
- Darst, R. P., Wang, D. & Auble, D. T. MOT1-catalyzed TBP–DNA disruption: uncoupling DNA conformational change and role of upstream DNA. *EMBO J.* **20**, 2028–2040 (2001).
- Sprouse, R. O., Brenowitz, M. & Auble, D. T. Snf2/Swi2-related ATPase Mot1 drives displacement of TATA-binding protein by gripping DNA. *EMBO J.* **25**, 1492–1504 (2006).
- Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature Struct. Biol.* **13**, 603–610 (2006).
- Gangaraju, V. K., Prasad, P., Srour, A., Kagalwala, M. N. & Bartholomew, B. Conformational changes associated with template commitment in ATP-dependent chromatin remodeling by ISW2. *Mol. Cell* **35**, 58–69 (2009).
- Dasgupta, A., Darst, R. P., Martin, K. J., Afshari, C. A. & Auble, D. T. Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. *Proc. Natl Acad. Sci. USA* **99**, 2666–2671 (2002).
- Kleiman, M. P. *et al.* NC2 $\alpha$  interacts with BTA1 and stimulates its ATP-dependent association with TATA-binding protein. *Mol. Cell Biol.* **24**, 10072–10082 (2004).
- van Werven, F. J. *et al.* Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. *Genes Dev.* **22**, 2359–2369 (2008).
- Hsu, J. Y. *et al.* TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes Dev.* **22**, 2353–2358 (2008).
- Geisberg, J. V. & Struhl, K. Cellular stress alters the transcriptional properties of promoter-bound Mot1–TBP complexes. *Mol. Cell* **14**, 479–489 (2004).
- Kostrewa, D. *et al.* RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
- Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010).
- Schluesche, P., Stelzer, G., Piaia, E., Lamb, D. C. & Meisterernst, M. NC2 mobilizes TBP on core promoter TATA boxes. *Nature Struct. Biol.* **14**, 1196–1201 (2007).
- Timmers, H. T., Meyers, R. E. & Sharp, P. A. Composition of transcription factor B-TFIID. *Proc. Natl Acad. Sci. USA* **89**, 8140–8144 (1992).
- Poon, D., Campbell, A. M., Bai, Y. & Weil, P. A. Yeast Taf170 is encoded by MOT1 and exists in a TATA box-binding protein (TBP)–TBP-associated factor complex distinct from transcription factor IID. *J. Biol. Chem.* **269**, 23135–23140 (1994).
- Gkikopoulos, T., Havas, K. M., Dewar, H. & Owen-Hughes, T. SWI/SNF and Asf1p cooperate to displace histones during induction of the *Saccharomyces cerevisiae* HO promoter. *Mol. Cell Biol.* **29**, 4057–4066 (2009).
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
- Auble, D. T., Wang, D., Post, K. W. & Hahn, S. Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. *Mol. Cell Biol.* **17**, 4842–4851 (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the Max-Planck Crystallization Facility Martinsried. We thank M. Lucas, A. Schele, C. Ungewickell, J. Goetzl and Y. Hiruma for help with experimentation. We thank J.-P. Armache and M. Turk for help with electron microscopy data. We are grateful to G. Miller and S. Hahn for advice. We thank the staff at the SLS and ESRF for help with data collection. We thank P. Cramer and members of the Hopfner and Auble laboratories for discussions and comments on the manuscript. This work was supported by the German Research Council (SFB 646 and SFB/TR5) and Excellence Initiative (Center for Integrated Protein Science, Munich) to K.-P.H. and R.B., by DFG grant WE4628/1 to P. Wendler and by NIH grant GM55763 to D.T.A.

**Author Contributions** S.C. and M.M. cloned, purified and crystallized EcTBP; S.C. solved its structure. S.C., A.B., M.M. and P. Wollmann cloned, purified and crystallized EcTBP–EcMot1(NTD); P. Wollmann collected data and P. Wollmann, G.W. and K.-P.H. solved the complex structures. R.V. performed FeBABE experiments, M.N.W. conducted yeast molecular biological manipulations and D.T.A. performed gelshifts. P. Wendler, O.B. and R.B. performed and interpreted electron microscopy experiments. P. Wollmann, P. Wendler, R.B., D.T.A. and K.-P.H. planned and interpreted the experiments. D.T.A. and K.-P.H. wrote the manuscript and all authors provided editorial input.

**Author Information** Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession codes 3OC1 (EcTBP) and 3OC3 (EcTBP–EcMot1(NTD) complex). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.-P.H. ([hopfner@lmb.uni-muenchen.de](mailto:hopfner@lmb.uni-muenchen.de)) or D.T.A. ([auble@virginia.edu](mailto:auble@virginia.edu)).

## METHODS

**Protein preparation.** BL21 Rosetta *E. coli* cells (Novagen) were used for expressing EcTBP and EcTBP(K103E) (pET28, Novagen), and for co-expressing EcTBP and EcMot1(NTD) (residues 1–778) or EcMot1(NTD $\Delta$ latch) (pET-DUET, Novagen). Proteins were purified by Ni<sup>2+</sup>-affinity chromatography (Qiagen) using a high-salt buffer at pH 8. Further purification of EcTBP was achieved by anion exchange chromatography (HiTrap SP HP, GE Healthcare). For crystallization of EcTBP, the His-tag was removed by tobacco etch virus (TEV) protease digestion. Final purification of all proteins was performed by size exclusion chromatography (Superdex S200, GE Healthcare). Production of selenomethionine-labelled EcTBP and EcMot1(NTD) was done in *E. coli*. Purification of EcTBP and EcMot1(NTD $\Delta$ latch), or of selenomethionine-labelled EcTBP and EcMot1(NTD), was performed accordingly.

Sequences encoding full-length EcMot1 (residues 1–1275), EcMot1( $\Delta$ CT) (residues 1–1016), EcMot1(NTD) (residues 1–779) and EcMot1( $\Delta$ latch) ( $\Delta$ 96–132), including an N-terminal 10  $\times$  His-tag, were cloned into the pFBDM transfer vector (Invitrogen). EcMot1(E912Q) (Walker B mutant) was generated by site-directed mutagenesis of pFBDM-EcMot1. Transposition of the coding sequence into MultiBac baculoviral DNA was performed in *E. coli* DH10MultiBac<sup>Cre</sup> cells<sup>31</sup>. Isolated bacmid DNA was used for transfection of *Trichoplusia ni* High Five insect cells (Invitrogen) to produce baculovirus for large-scale infections. Proteins were purified by Ni<sup>2+</sup>-affinity chromatography (Qiagen) using buffer containing 50 mM MES (pH 6.5), 200 mM NaCl, 10 mM  $\beta$ -mercaptoethanol and 12.5 mM or 300 mM imidazole (for EcMot1(NTD) and EcMot1(NTD $\Delta$ latch)), or 50 mM Tris (pH 7.5), 400 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 10% glycerol (v/v) and 12.5 mM or 300 mM imidazole (for EcMot1, EcMot1( $\Delta$ CT) and EcMot1( $\Delta$ latch)). Additional purification was achieved by ion exchange chromatography (HiTrap Q HP, GE Healthcare). For crystallization of the complex, EcTBP was added in excess amounts to EcMot1. Final purification of the proteins was done by size exclusion chromatography (Superdex S200, GE Healthcare). The preparation of EcMot1(BeF)<sub>3</sub>-EcTBP (EcMot1 with ADP-BeF<sub>3</sub><sup>−</sup>) was performed as described previously<sup>32</sup>.

**Crystallization.** Proteins were crystallized by hanging-drop vapour diffusion at 18 °C in a mixture of 1  $\mu$ l protein (10 mg ml<sup>−1</sup> EcTBP and 5 mg ml<sup>−1</sup> EcMot1(NTD)-EcTBP) and 1  $\mu$ l precipitant (0.1 M 2-(*N*-morpholino)ethane sulphonic acid (pH 6.5), 2 M NaCl and 4% acetone for EcTBP; 50 mM MES (pH 6), 200 mM ammonium acetate, 5% 2-methyl-2,4-pentanediol, 4% polyethylene glycol 3350 and 200 mM 3-(1-pyridino)-1-propane sulphate (NDSB-201) for EcMot1(NTD)-EcTBP). Crystals were cryoprotected with 1,2-ethanediol (EcTBP) or 2,3-butanediol (EcMot1(NTD)-EcTBP) and flash-frozen in liquid nitrogen.

**Structure determination.** EcTBP crystals diffracted to 1.9 Å resolution and were measured at the Swiss Light Source (SLS). Native data of crystals from EcMot1(NTD)-EcTBP showed that they diffracted X-rays to 3.1 Å; these data were collected at the European Synchrotron Radiation Facility (ESRF). Data from derivative crystals of selenomethionine-labelled EcMot1(NTD)-EcTBP were collected to 3.3 Å at the SLS. All data were processed with XDS<sup>33</sup>. The structure of EcTBP was solved by molecular replacement using Phaser<sup>34</sup> and yeast TBP (PDB code 1TBP) as a model. The structure of EcMot1(NTD)-EcTBP was solved by a single anomalous dispersion experiment using SeMet data in combination with molecular replacement, using the EcTBP structure as a partial model (Phaser<sup>34</sup>). Heavy-atom sites were obtained with SHARP<sup>35</sup> and initial automatic model building was performed with Buccaneer<sup>36</sup>. Model building and refinement was conducted in Coot<sup>37</sup> and PHENIX<sup>38</sup>, respectively. Figures were prepared in Pymol<sup>39</sup> or Chimera<sup>40</sup>.

**Footprinting assays.** Footprinting assays were performed as described previously<sup>41</sup>. Reactions contained 20 nM EcTBP, 30 nM EcMot1 and 50 mM ATP, as indicated. After incubation of TBP with DNA for 20 min at 37 °C, Mot1 was added with or without ATP for 5 min before DNase I digestion and sample processing.

**ATPase assay.** The rates of ATP hydrolysis were measured as described previously<sup>41</sup> in buffer containing 4 mM Tris-HCl (pH 8), 60 mM KCl, 5 mM MgCl<sub>2</sub>, 4% glycerol (v/v), 100 mg ml<sup>−1</sup> BSA and 1 mM dithiothreitol at 22 °C.

**Analytical gel filtration.** Analytical size exclusion experiments were performed on Ettan LC system (GE Healthcare, Superose 12 PC 3.2/30) according to the manufacturer's instructions (50 mM HEPES (pH 8) or 50 mM MES (pH 6.5), 200 mM NaCl and 2 mM dithiothreitol).

**Electrophoretic mobility shift assays.** These assays used a radiolabelled fragment of the adenovirus major late promoter<sup>41</sup>. Typically, <1 nM DNA was incubated with 15–20 nM EcTBP for 20 min at 37 °C in the same buffer as was used for ATPase assays, then 30 nM Mot1 (or a Mot1 mutant) was added with or without 50 mM ATP for 5 min, before loading on a gel as previously described<sup>42</sup>.

**Dissociation kinetic assays.** Kinetic analysis of the dissociation reaction was performed by addition of ATP to pre-formed ternary complexes under the

conditions used in the experiment shown in Fig. 2e. Reactions contained radio-labelled DNA (<1 nM), 20 nM EcTBP and either 30 nM wild-type EcMot1 or 30 nM EcMot1( $\Delta$ latch). EcTBP was incubated with the radiolabelled DNA template for 20 min, followed by addition of EcMot1 or EcMot1( $\Delta$ latch) for 10 min. ATP was added to 100  $\mu$ M for 2–20 min and reaction products were resolved at the indicated times on non-denaturing gels. To quantify the extent of complex dissociation at each time point, the free DNA band was quantified and expressed as a proportion of the free DNA present in reactions with no added protein. The results are expressed as the average  $\pm$  standard error associated with two independent experiments.

**FeBABE cleavage assays.** FeBABE (Dojindo) was conjugated to 68-bp DNA duplexes, based on the sequence of the adenovirus major late promoter. Biotinylation of the top strand's 5' end allowed the duplexes to be bound by streptavidin beads. After FeBABE conjugation, TBP and Mot1 were loaded onto the modified DNAs and cutting was initiated by addition of ascorbic acid and hydrogen peroxide. FeBABE-mediated protein cleavage has been previously described<sup>43,44</sup>. The yeast system was used to take advantage of an antibody raised to the C terminus of yeast Mot1 (ref. 45).

**Static light scattering.** For molecular weight determination of protein samples (2–4 mg ml<sup>−1</sup>, with 40 mM HEPES (pH 8), 200 mM NaCl and 2 mM dithiothreitol as a running buffer) by static light scattering, we used a combination of a Viscotek 270 detector and a Viscotek VE-3580 refractive index monitor connected to a microscale HPLC system (AEKTAmicro, GE Healthcare) equipped with an analytical size exclusion column (Superdex S200 15/150 GL, GE Healthcare). Data analysis was performed using the OmniSEC software (Viscotek) using BSA (Thermo Fisher) as a reference for calibration. The chromatographs of the size exclusion, monitored by ultraviolet absorption at 280 nm, and the subsequent refractive index and light-scattering chromatographs all showed a single prominent peak indicating a homogenous sample. Plots of the determined molecular weight versus elution volumes for the evaluated peaks all showed stable molecular weights for the chosen peak areas.

**Dynamic light scattering.** Dynamic light scattering was measured using a Viscotek/Malvern Instruments 802DLS system. Protein samples (1 mg ml<sup>−1</sup> in size exclusion buffer) were centrifuged and the supernatant was measured at 20 °C using fluorescence cuvettes. At least ten autocorrelation curves per sample were recorded, averaged and evaluated using the OmniSIZE software and the mass model for globular proteins. All samples showed intensity distributions indicating a homogenous sample with a single peak at the given hydrodynamic radius.

**Electron microscopy.** 3.5  $\mu$ l (10–30  $\mu$ g ml<sup>−1</sup>) of freshly prepared protein sample was applied to pre-coated Quantifoil holey carbon-supported grids and negatively stained using 2% uranyl acetate. Micrographs were recorded on a Tecnai G2 Spirit TEM at 120 kV. Data were collected under low-dose conditions at a nominal magnification of  $\times 90,000$  and a nominal defocus of  $-0.9 \mu$ m using an Eagle 2048  $\times$  2048 pixel CCD camera (FEI Company) with a resolution of 30  $\mu$ m pixel<sup>−1</sup> (3.31 Å pixel<sup>−1</sup> object scale). 5,518 particles of EcMot1(BeF)<sub>3</sub>-EcTBP (EcMot1-EcTBP in the presence of 2 mM ADP and beryllium fluoride), 7,737 of EcMot1( $\Delta$ CT)-EcTBP and 12,558 of EcMot1(E912Q) (Walker B mutant of EcMot1 was used instead of wild type owing to its enhanced stability) were picked using boxer<sup>46</sup>. Initial image processing was done using IMAGIC-5 (ref. 47). The images were normalized, filtered at the first zero without CTF correction and centred by iteratively aligning them to their rotationally averaged sum. Initial class averages were obtained by 2–3 rounds of multivariate statistical analysis, followed by multi-reference alignment using homogenous classes as references. The data sets were classified into

10–20 images per class. A low-resolution density map was created by angular reconstruction and was used as an initial model for projection-matching in EMAN 1.9 (ref. 46). The models underwent 8–24 rounds of refinement at an angular increment of up to 5 degrees, until angular assignment was stable. The final reconstructions comprised approximately 90% of the original data set. All visualization and rigid-body fittings were carried out using the UCSF Chimera package<sup>40</sup>. Surface representations show density rendered at a threshold accounting for the expected molecular mass of the complexes: EcMot1( $\Delta$ CT)-EcTBP (140 kDa; 170,226 Å<sup>3</sup>), EcMot1(E912Q) (145 kDa; 172,366 Å<sup>3</sup>) and EcMot1(BeF)<sub>3</sub>-EcTBP (169 kDa; 198,984 Å<sup>3</sup>). Crystal structures and density maps were merged in VMD using MDFF, as described previously<sup>48</sup>. Difference densities were created by subtracting from one another maps that were rendered at a threshold accounting for the expected molecular mass of the complex.

- Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnol.* **22**, 1583–1587 (2004).
- Kagawa, R., Montgomery, M. G., Braig, K., Leslie, A. G. & Walker, J. E. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. *EMBO J.* **23**, 2734–2744 (2004).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).

34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
35. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
36. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
38. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
39. DeLano, W. L. *The PyMOL Molecular Graphics System* Version 1.3 r1 (Schrödinger, 2010).
40. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
41. Sprouse, R. O., Brenowitz, M. & Auble, D. T. Snf2/Swi2-related ATPase Mot1 drives displacement of TATA-binding protein by gripping DNA. *EMBO J.* **25**, 1492–1504 (2006).
42. Darst, R. P., Wang, D. & Auble, D. T. MOT1-catalyzed TBP-DNA disruption: uncoupling DNA conformational change and role of upstream DNA. *EMBO J.* **20**, 2028–2040 (2001).
43. Chen, H. T. & Hahn, S. Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol. Cell* **12**, 437–447 (2003).
44. Miller, G. & Hahn, S. A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nature Struct. Mol. Biol.* **13**, 603–610 (2006).
45. Auble, D. T., Wang, D., Post, K. W. & Hahn, S. Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. *Mol. Cell. Biol.* **17**, 4842–4851 (1997).
46. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
47. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
48. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).



# Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF

Cameron D. Mackereth<sup>1,2,3</sup>, Tobias Madl<sup>1,4</sup>, Sophie Bonnal<sup>5</sup>, Bernd Simon<sup>3</sup>, Katia Zanier<sup>3</sup>, Alexander Gasch<sup>3</sup>, Vladimir Rybin<sup>3</sup>, Juan Valcárcel<sup>5,6</sup> & Michael Sattler<sup>1,3,4</sup>

Many cellular functions involve multi-domain proteins, which are composed of structurally independent modules connected by flexible linkers. Although it is often well understood how a given domain recognizes a cognate oligonucleotide or peptide motif, the dynamic interaction of multiple domains in the recognition of these ligands remains to be characterized. Here we have studied the molecular mechanisms of the recognition of the 3'-splice-site-associated polypyrimidine tract RNA by the large subunit of the human U2 snRNP auxiliary factor (U2AF65)<sup>1–3</sup> as a key early step in pre-mRNA splicing<sup>4</sup>. We show that the tandem RNA recognition motif domains of U2AF65 adopt two remarkably distinct domain arrangements in the absence or presence of a strong (that is, high affinity) polypyrimidine tract. Recognition of sequence variations in the polypyrimidine tract RNA involves a population shift between these closed and open conformations. The equilibrium between the two conformations functions as a molecular rheostat that quantitatively correlates the natural variations in polypyrimidine tract nucleotide composition, length and functional strength to the efficiency to recruit U2 snRNP to the intron during spliceosome assembly<sup>1,5–8</sup>. Mutations that shift the conformational equilibrium without directly affecting RNA binding modulate splicing activity accordingly. Similar mechanisms of cooperative multi-domain conformational selection may operate more generally in the recognition of degenerate nucleotide or amino acid motifs by multi-domain proteins<sup>9,10</sup>.

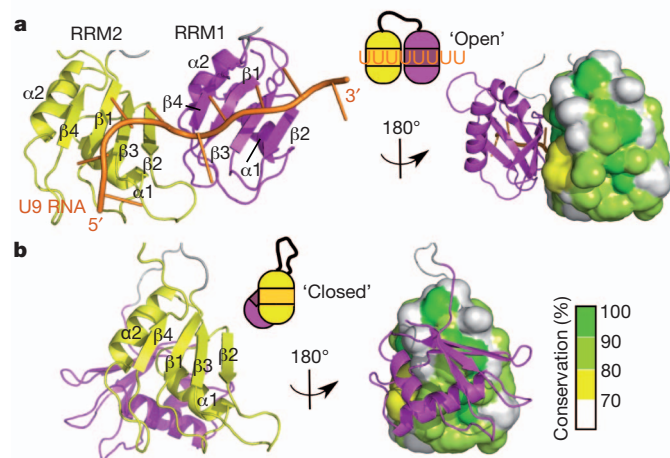
The essential multi-domain splicing factor U2AF65 has a crucial role in the assembly of splicing complexes<sup>4</sup>. A polypyrimidine (Py) tract RNA sequence at the 3' end of introns is recognized by the tandem RNA recognition motif (RRM) domains (RRM1–RRM2) of U2AF65 (refs 11, 12). However, there is significant diversity in the nucleotide composition, length and functional strength of the Py tract sequence, reflecting the dynamic range of splice site acceptor site usage in events such as alternative splicing. The mechanisms by which the Py tract sequence variations found in human U2 introns<sup>1,5–8</sup> are recognized by U2AF65 and how the 'strength' of a given Py tract is coupled to the efficiency of spliceosome assembly are not understood. Using a novel protocol for structural analysis of multi-domain proteins and protein complexes in solution<sup>13</sup> (Supplementary Text and Methods), we studied the minimal region in U2AF (U2AF65 RRM1–RRM2, residues 148–342; Supplementary Fig. 1a) that mediates binding to the Py tract RNA and recapitulates the key features of Py tract recognition by U2AF (Supplementary Text and Supplementary Figs 2–4).

We found that the U2AF65 RRM1–RRM2 tandem domains can populate two distinct three-dimensional arrangements correlated to the presence or absence of a high-affinity RNA ligand (Fig. 1a, b and Supplementary Tables 1 and 2). In the 'open' conformation of the RRM1–RRM2 tandem domains, as observed when bound to U9 RNA (Fig. 1a), a parallel arrangement of the two  $\beta$ -sheets forms an extended basic RNA-binding surface (Supplementary Fig. 5). The protein–protein

interface between the two RRMs involves residues from  $\alpha 2$  to  $\beta 4$  in RRM1 and  $\alpha 1$ ,  $\beta 2$  and the  $\beta 2$ – $\beta 3$  linker in RRM2, stabilized mainly through electrostatic complementarity. The RRM1–RRM2–U9 model also incorporates atomic details of protein–RNA contacts for the individual RRMs seen in the previous crystal structure<sup>12</sup> (Supplementary Fig. 6). However, the nuclear magnetic resonance (NMR) data are inconsistent with the overall arrangement of the tandem RRM domains in the crystal (Supplementary Fig. 7), indicating that the relative domain orientation was influenced by crystal packing forces and/or deletion of the linker (which is conserved in length, Supplementary Fig. 8).

In the second, 'closed' conformation, observed in the absence of ligand (see Supplementary Text for details on structure calculation), the RNA-binding surface of RRM1 ( $\beta 2$ ) is partially occluded by an interaction with helices  $\alpha 1$  and  $\alpha 2$  of RRM2 (Fig. 1b). The protein–protein interface between RRM1 and RRM2 in this 'closed' conformation agrees well with residues identified based on chemical shift differences between RRM1–RRM2 and the isolated domains (Supplementary Fig. 9c, d) and is further supported by an excluded solvent-accessible area (derived from solvent paramagnetic relaxation enhancement (PRE) data; not shown). As in the RNA-bound form, the domain interface comprises mainly electrostatic interactions involving conserved residues (Supplementary Fig. 8).

One set of measurements required for model generation involves long-range distance restraints derived from PRE. PREs are obtained by spin labelling various residues in RRM1–RRM2 and are detected as



**Figure 1 | Structure of the tandem RRM domains of U2AF65 free and when bound to a high-affinity Py tract.** **a, b,** Cartoon and ribbon representation of the lowest energy solution structure models calculated for the (a) RNA-bound or open form of RRM1–RRM2 with a U9 Py tract RNA (orange), and (b) the unbound or closed form of RRM1–RRM2. The conserved surface of RRM2 is exposed in the open conformation (a, right) but is occluded by RRM1 (shown as magenta ribbon) in the free protein or in the presence of weak Py tracts (b, right).

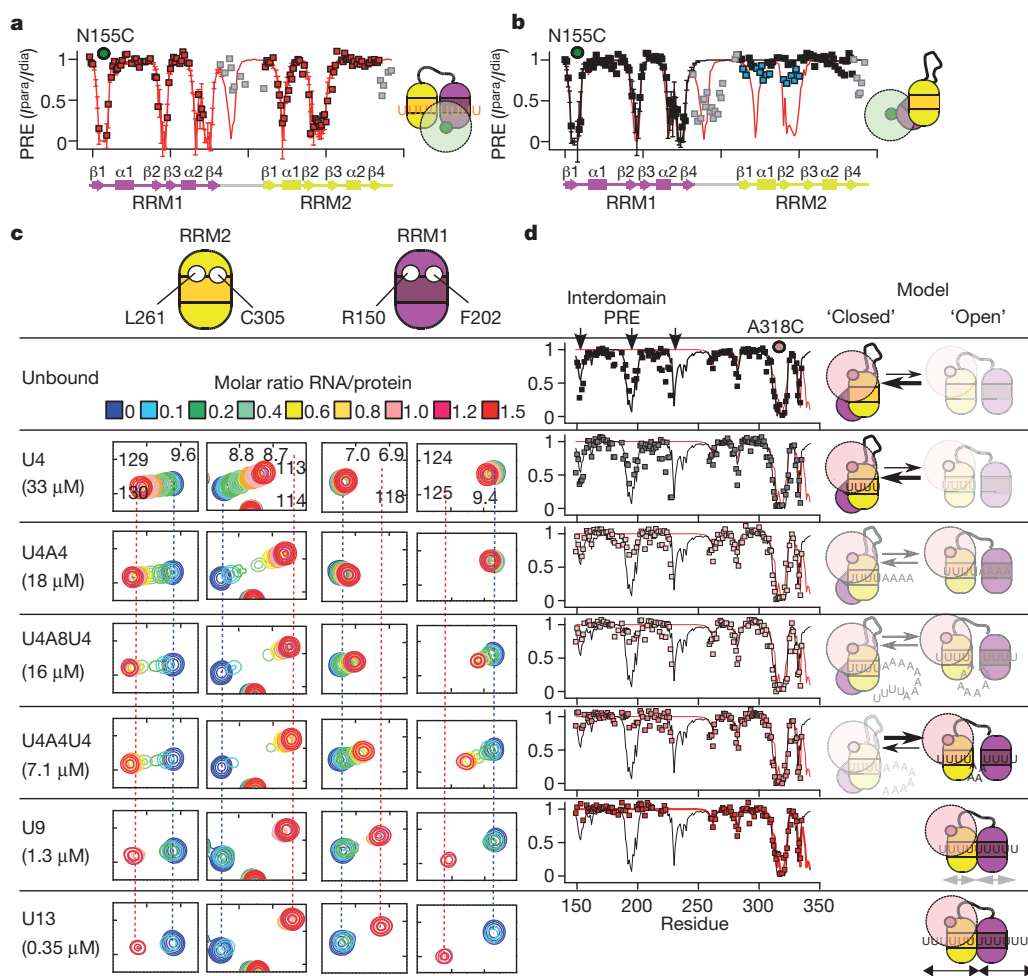
<sup>1</sup>Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. <sup>2</sup>Institut Européen de Chimie et Biologie and Université de Bordeaux, 2 rue Robert Escarpit, 33607 Pessac, France. <sup>3</sup>Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. <sup>4</sup>Munich Center for Integrated Protein Science and Chair Biomolecular NMR, Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany. <sup>5</sup>Centre de Regulació Genòmica, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain. <sup>6</sup>Institució Catalana de Recerca i Estudis Avançats, Dr. Aiguader 88, 08003 Barcelona, Spain.

line-broadening of the NMR signals depending on the distance from the spin label (see for example Fig. 2a, b, d for spin labels at residues 155 and 318, respectively). Notably, the PRE data for the RNA-free sample indicate the presence of a pre-existing, minor population of RRM1–RRM2 that corresponds to the open form (blue squares in Fig. 2b and Supplementary Fig. 9b). This indicates that conformational sub-states resembling the open conformation of RRM1–RRM2 exist already in the absence of RNA ligand. Although similar observations have been described for other systems<sup>14–19</sup>, an equilibrium between two distinct open and closed states for binding of multi-domain proteins to a degenerate ligand motif has not been reported. We thus wondered whether Py tract recognition by U2AF65 may involve a graded shift in a pre-existing ‘multi-domain’ equilibrium by conformational selection and thus provide the molecular rationale linking the wide variety of intron RNA Py tract sequences to their encoded ‘strength’ of splicing efficiency.

To this end, we examined whether a dynamic equilibrium between the open and closed forms of U2AF65 RRM1–RRM2 could provide a mechanism for regulating the extent of Py tract binding. In the absence of RNA (closed conformation) only the RNA-binding surface of RRM2 is freely accessible for initial interactions with RNA. Thus, short RNA ligands that can cover only a single RRM domain (such as a four-uridine RNA) should bind preferentially to RRM2 and fail to alter the domain rearrangement. Consistent with this, titration of RRM1–RRM2 with U4 RNA shows significant chemical shift perturbations

only for residues in RRM2 (Fig. 2c) and the binding affinity of U4 to RRM1–RRM2 is comparable to the interaction with the isolated RRM2 (Supplementary Table 3). Moreover, the pattern of inter-domain PRE data and therefore the relative domain arrangement is very similar to that of the unbound RRM1–RRM2 (Fig. 2d and Supplementary Fig. 10). This indicates that U4 mainly binds to RRM2, and that RRM1–RRM2 is predominantly in the ‘closed’ conformation when bound to U4.

We then investigated the conformation of RRM1–RRM2 upon binding to a series of RNA ligands, representing Py tracts of various length and composition, mimicking the degeneracy of Py tracts found in human U2 introns<sup>1,5–8</sup>. Using isothermal titration calorimetry (ITC), NMR chemical shift perturbation and PRE measurements, we found that the ligands U4A4, U4A8U4 and U4A4U4 show intermediate but gradually increasing affinity to RRM1–RRM2, in comparison to the low-affinity U4 and the high-affinity U9/U13 ligands (Fig. 2c, d). Surprisingly, each Py tract RNA shows similar binding to RRM2 regardless of the overall binding affinity (Fig. 2c; comparable chemical shift perturbation for RRM2 residues). Instead, the overall increase in affinity reflects an increasing contribution of RRM1 bound to RNA, as shown by the extent of chemical shift perturbation seen for residues in RRM1 (Fig. 2c). Full binding of RRM1 appears only with a long uninterrupted stretch of pyrimidine, such as with U9 or U13, with adjacent high-affinity binding sites for both RRM1 and RRM2, and an overall affinity approaching the product of the U4 RNA affinities



**Figure 2 | Binding of Py tracts of different strength to U2AF65 RRM1–RRM2.** **a, b,** Paramagnetic relaxation enhancement (PRE) data from a spin label attached to residue 155 (green circle) for **(a)** U9-bound RRM1–RRM2 (red squares) and **(b)** unbound RRM1–RRM2 (black squares), with back-calculated PRE values (red line, U9 bound; black line, unbound) and derived from the structure ensembles (s.d. from mean). Flexible regions and minor open population data are shown by grey and blue squares, respectively. **c,** RRM1–

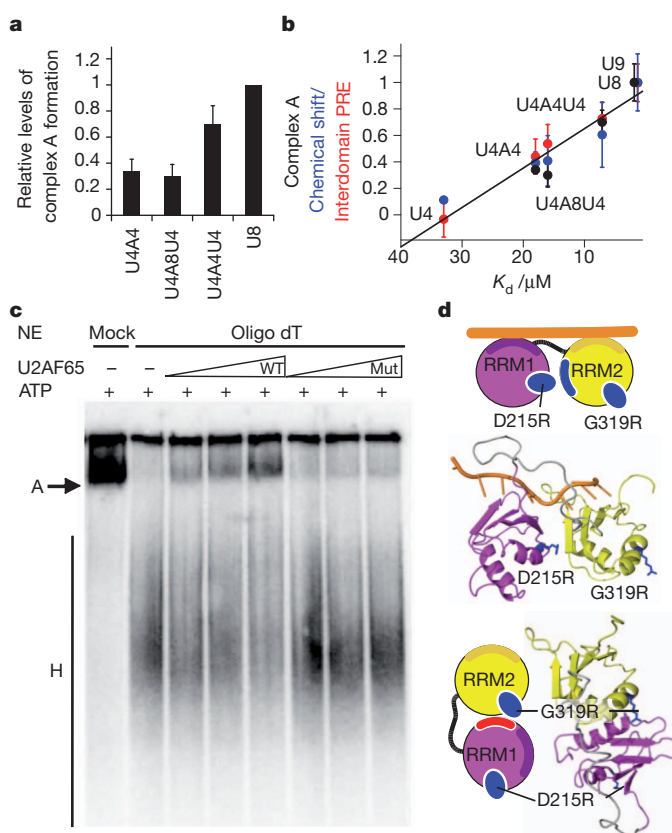
RRM2 titration NMR spectra with model Py tracts (dissociation constants from Supplementary Table 3 in parentheses);  $^1\text{H}/^{15}\text{N}$  chemical shifts in parts per million are shown on the  $x$ - $y$ -axis. **d,** Experimental PRE data (peak intensities in the paramagnetic and diamagnetic state ( $I^{\text{para}}/I^{\text{dia}}$ ) versus residue number) and back-calculations for unbound and the various RRM1–RRM2–RNA ligand complexes for a spin label attached to residue 318 (pink circle) and a corresponding schematic of the equilibrium between open and closed states.

of the individual domains. Analysis of the PRE data shows a gradual change in the pattern of inter-domain PRE from the unbound form (and U4-bound) to the fully bound form (Fig. 2d and Supplementary Fig. 10a). These complexes are still mainly composed of compact states, where the two domains interact and reorient together in solution, as shown by NMR relaxation data (Supplementary Fig. 10b). As the NMR data report on population-weighted averages of the molecules in solution, it is reasonable to assume that RRM1–RRM2 exists in equilibrium between the two conformations, corresponding to the open (bound) and closed (unbound) RRM1–RRM2 structures, respectively. Therefore, the data shown in Fig. 2 indicate that binding of Py tracts of increasing affinity (or strength) results in a shift of populations between the closed and open conformations (Supplementary Fig. 10c–e). More importantly, the data indicate that RRM1 is a key regulator of this mechanism governed by the competition between binding RRM2 and binding a secondary RNA site within the Py tract.

To analyse whether this population shift from the closed to open conformation is coupled to the functional U2AF activity during spliceosome assembly, we measured U2 snRNP recruitment (pre-spliceosome (A) complex formation) on RNAs containing the 3'-splice-site region and downstream exon of adenovirus major late (AdML) promoter transcripts with both native (U8) and the different Py tract configurations analysed above (U4A4, U4A8U4, U4A4U4). Complex A is formed most efficiently with the U8 Py tract, somewhat less with U4A4U4, and with significantly lower efficiency using the U4A8U4 and U4A4 substrates (Fig. 3a and Supplementary Fig. 11a, b). There is a notable quantitative correlation between the extent of U2 snRNP recruitment, RNA binding affinity and the population of molecules adopting the open conformation of U2AF65 (Fig. 3b and Supplementary Fig. 11c). The similarity between the U4A8U4 and U4A4 substrates suggests that an eight-adenosine spacing between two consecutive uridine stretches is unable to compete for RRM1 binding against the RRM1–RRM2 interaction present in the closed conformation. Notably, the results obtained using the model RNA ligands also extend to native Py tracts represented by four human intron sequences that contain comparable length and branch-point strength (Supplementary Fig. 12).

We next designed mutants of RRM1–RRM2 with the aim to shift the equilibrium between open and closed states, thereby perturbing the degree of conformational sampling and thus affecting the formation of complex A accordingly. Several mutations were created remote from the RNA-binding surface (Supplementary Fig. 13) and investigated using ITC (Supplementary Table 3). The double mutation D215R/G319R destabilizes the open conformation (D215R) by electrostatic repulsion, whereas it strengthens the interface of the closed conformation with favourable charge complementarity (G319R) (Fig. 3d). This mutant shows reduced affinity for U9 and U4A8U4 consistent with a shift of the conformational equilibrium towards the closed state (Supplementary Fig. 14) and shows strongly reduced formation of complex A (Fig. 3c). A second mutant, RRM1–RRM2( $\Delta$ 233–252), was designed to selectively prevent the closed conformation due to a strategic shortening of the linker connecting RRM1 and RRM2. As predicted, RRM1–RRM2( $\Delta$ 233–252) favours the open conformation even in the absence of RNA ligand, as confirmed by the pattern of PRE (Supplementary Fig. 15). In addition, this mutant has increased binding to U9, U4A4U4 and U4A8U4 ligands. It displays a consistent level of binding by RRM1 in keeping with removal of the competition between binding RRM2 and RNA and shows activity in splicing assays comparable to the wild-type protein (Supplementary Fig. 15). The predicted opposing effects of these two mutants provide further support for the functional significance of the conformational equilibrium of the U2AF65 tandem RRM domains.

Our results indicate that the tandem RRM domains of U2AF65 do not simply act as a binding scaffold but instead have an active role in quantitatively relating Py tract strength to splice site recognition and spliceosome assembly (Fig. 4 and Supplementary Fig. 16). Multi-domain



**Figure 3 | Spliceosome assembly as a function of Py tract strength.**

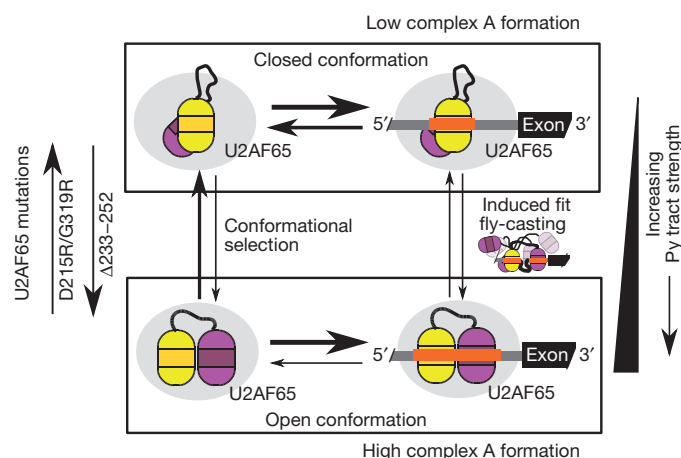
**a**, Complex A formation for AdML promoter transcripts with various Py tract sequences (Supplementary Fig. 10). Error bars indicate mean  $\pm$  s.d. for 11 replicates. **b**, Correlation of binding affinity with: complex A formation (black; from **a**); relative inter-domain PRE effect from the open conformation population (red; spin label at 318, Fig. 2d; error bars from 100 iterations of a Monte Carlo analysis) and relative average chemical shift perturbation of RRM1 (blue), with error bars for mean  $\pm$  s.d. The black line represents a linear fit of the data. **c**, Complex A formation in U2AF-depleted nuclear extracts with recombinant purified GST–U2AF65 (WT) or mutant D215R/G319R (Mut). Spliceosomal complexes A and H are indicated by 'A' and 'H' on the left. **d**, Design rationale for the D215R/G319R mutant.

conformational selection of the open states allows the tandem RRM domains to function as a molecular rheostat with regard to U2AF activity during early steps of splicing, involving a competition for RRM1 between binding RRM2 (autoinhibition in a closed conformation) and RNA (activation by an open conformation). This provides a selectivity filter against promiscuous RNA binding and spliceosome assembly, as the higher affinity Py tract ligands are better able to counteract the energetic penalty needed for both RRM domains to bind.

Our data do not rule out the existence of a minor induced fit mechanism involving 'fly-casting'<sup>20</sup> where the tandem RRM domains may be able to identify weak Py tracts where short pyrimidine stretches are distributed over a longer RNA sequence. After initial binding of a short (that is, four-nucleotide) pyrimidine stretch to RRM2, neighbouring uridine stretches can be screened by RRM1 to find a complete 8-mer Py tract with increased U2AF affinity. The search space of RRM1 is restricted by the conserved length (not the sequence) of the linker connecting RRM1 and RRM2 (Supplementary Fig. 8). In addition, depending on the separation of the two U4 stretches, the entropy loss associated with binding of the RNA to RRM1–RRM2 decreases as the RNA linker is shortened. This will affect the relative contribution of induced fit compared to conformational selection (Supplementary Fig. 16)<sup>21</sup>.

The tunable conformational shift described here can contribute to overall 3'-splice-site recognition beyond simply improving U2AF RNA occupancy. The open conformation may expose protein functionalities that are occluded in the closed RRM1–RRM2 state, and may thereby





**Figure 4 | Py tract recognition by U2AF.** A multi-domain conformational-selection mechanism enables Py tracts of increasing strength to capture the open conformation of U2AF and support efficient assembly of complex A. Protein mutations can shift the equilibrium to favour either the open or closed conformation (left). Relative sizes of the on- and off-rates are indicated by the thickness of the arrows (see Supplementary Text and Supplementary Fig. 16 for further details). Fly-casting may represent a minor mechanism of induced fit based on the extent of spatial separation of Py tract elements.

facilitate U2 snRNP recruitment. Most notable is the conserved  $\alpha$ -helical surface of RRM2 that is only accessible in the open orientation (Fig. 1a, b). This region contains a lysine residue (K276) that upon hydroxylation alters the splicing pattern of some genes<sup>22</sup>. The equilibrium between open and closed conformations might therefore orchestrate a distinct ribonucleoprotein assembly characteristic of activated 3' splice sites. Reciprocally, additional protein-binding partners of U2AF65 (for example, U2AF35 or others<sup>23</sup>), through additional interactions with 3'-splice-site components (for example, the AG dinucleotide), could favour the open conformation and thereby enhance the recognition of weak Py tracts.

We expect that similar mechanisms of multi-domain conformational selection coupled to biological activity operate in many multi-domain proteins that must functionally distinguish degenerate nucleotide or amino acid motifs from similar, nonspecific sequences. As demonstrated here for U2AF65, structural analysis of multiple domains connected by flexible linkers critically depends on the use of solution techniques in a multidisciplinary approach.

## METHODS SUMMARY

Wild-type and mutated U2AF65 constructs were cloned, expressed in *Escherichia coli* and purified as described in Methods. Oligoribonucleotides were purchased from Biospring GmbH. NMR spectra were collected at 295 K, with chemical shifts assigned by standard experiments or by comparison to previous data<sup>11</sup>. Residual dipolar coupling used partial alignment by Pf1 phage or a liquid crystal containing hexanol and pentaethylene glycol monododecyl ether<sup>24</sup>. NMR spectroscopy and structure calculation details are provided in Methods and Supplementary Information. Protein–RNA affinity was measured by isothermal titration calorimetry. *In vitro* assay of complex A assembly was carried out as described previously<sup>25</sup>; the splicing activity of U2AF65 mutants used recombinant protein and nuclear extracts, in which U2AF was depleted by oligo-dT cellulose chromatography<sup>26</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 9 July 2010; accepted 5 May 2011.

Published online 13 July 2011.

1. Zamore, P. D., Patton, J. G. & Green, M. R. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609–614 (1992).
2. Banerjee, H., Rahn, A., Davis, W. & Singh, R. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA* **9**, 88–99 (2003).
3. Banerjee, H. *et al.* The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF65) is essential *in vivo* but dispensable for activity *in vitro*. *RNA* **10**, 240–253 (2004).

4. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
5. Reed, R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.* **3**, 2113–2123 (1989).
6. Roscigno, R. F., Weiner, M. & Garcia-Blanco, M. A. A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J. Biol. Chem.* **268**, 11222–11229 (1993).
7. Singh, R., Valcárcel, J. & Green, M. R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**, 1173–1176 (1995).
8. Coolidge, C. J., Seely, R. J. & Patton, J. G. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**, 888–896 (1997).
9. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
10. Seet, B. T., Dikic, I., Zhou, M. M. & Pawson, T. Reading protein modifications with interaction domains. *Nature Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
11. Ito, T., Muto, Y., Green, M. R. & Yokoyama, S. Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)). *EMBO J.* **18**, 4523–4534 (1999).
12. Sickmier, E. A. *et al.* Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol. Cell* **23**, 49–59 (2006).
13. Simon, B., Madl, T., Mackereth, C. D., Nilges, M. & Sattler, M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew. Chem. Int. Edn Engl.* **49**, 1967–1970 (2010).
14. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5**, 789–796 (2009).
15. Zhang, Q., Stelzer, A. C., Fisher, C. K. & Al-Hashimi, H. M. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**, 1263–1267 (2007).
16. Lange, O. F. *et al.* Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).
17. Li, P., Martins, I. R., Amarasinghe, G. K. & Rosen, M. K. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nature Struct. Mol. Biol.* **15**, 613–618 (2008).
18. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fercht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329**, 1312–1316 (2010).
19. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916 (2007).
20. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl Acad. Sci. USA* **97**, 8868–8873 (2000).
21. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl Acad. Sci. USA* **106**, 13737–13741 (2009).
22. Webby, C. J. *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90–93 (2009).
23. Soares, L. M., Zanier, K., Mackereth, C., Sattler, M. & Valcárcel, J. Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* **312**, 1961–1965 (2006).
24. Rückert, M. & Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).
25. Guth, S., Tange, T., Ø., Kellenberger, E. & Valcárcel, J. Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol. Cell Biol.* **21**, 7673–7681 (2001).
26. Valcárcel, J., Martínez, C. & Green, M. R. Functional analysis of splicing factors and regulators. In *mRNA Formation and Function* 31–53 (Elsevier, 1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank F. Gabel, M. Nilges, C. Griesinger, J. Müller and K. Scheffzek for discussions, and H. Tilgner for analysis of natural Py tract sequences. C.D.M. acknowledges support by EMBO Long Term Fellowship, ICSN and Aquitaine regional government. T.M. thanks the Austrian Science Fund (FWF) and EMBO for postdoctoral fellowships. We thank the EU NMR LSF in Frankfurt and the Bavarian NMR Centre (BNMRZ) in Munich for NMR measurement time. This work was supported by the European Commission, grants 3D Repertoire, FSG-V-RNA and NIM3 No. 226507 (M.S.), EURASNET, AICR and Fundación Marcelino Botín (J.V.).

**Author Contributions** C.D.M., S.B., K.Z. and A.G. cloned and purified native and nitroxyl-labelled proteins. C.D.M., K.Z., B.S. and T.M. collected, processed and analysed NMR spectroscopy data. C.D.M., B.S. and T.M. calculated and analysed structural ensembles. S.B. performed *in vitro* splicing assays. V.R. performed ITC. J.V. and M.S. contributed to study design. C.D.M. and M.S. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** The coordinates of the open RNA-bound conformation of RRM1–RRM2 and the closed conformation in the absence of RNA are deposited in the Protein Data Bank with accession codes 2YH1 and 2YH0, respectively. All structural ensembles with explicit spin labels are available from the authors upon request. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.S. ([sattler@helmholtz-muenchen.de](mailto:sattler@helmholtz-muenchen.de)).

## METHODS

**Cloning.** Full-length human U2AF65, as well as the truncation mutants U2AF65(RRM1–RRM2), U2AF65(RRM1) and U2AF65(RRM2), were cloned by using PCR amplification. Primers were designed to introduce NcoI and Acc65I restriction enzyme sites, to allow for directional insertion into a modified pET9d vector containing an amino-terminal His<sub>6</sub> tag followed by a tobacco etch virus (TEV) protease cleavage site. Full-length U2AF65 constructs were cloned into a modified pET9d vector containing an N-terminal GST tag. The linker deletion and site-specific mutants were created by PCR amplification with overlapping oligonucleotides containing the mutated sequence. All plasmids were verified by sequencing.

**Expression and purification.** U2AF65-derived peptides were produced in BL21(DE3) or BL21(DE3)pLysS cells using standard media or minimal M9T media supplemented with 2 g l<sup>-1</sup> [<sup>13</sup>C]glucose and/or 1 g l<sup>-1</sup> [<sup>15</sup>N]ammonium chloride. Following normal growth, cells were induced at an OD<sub>600 nm</sub> of 0.6 with 0.25 μM IPTG followed by protein expression for 16 h at 25 °C. Cells were collected by centrifugation, lysed by sonication in the presence of lysozyme and EDTA-free Complete protease inhibitor (Roche Applied Science) then resuspended in binding buffer consisting of 50 mM Tris (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol and 5 mM imidazole. The sample was added to Ni<sup>2+</sup> affinity chromatography resin and washed with 20 column volumes of binding buffer followed by five column volumes of the same buffer but with 30 mM imidazole. Elution with 50 mM Tris (pH 7.5), 500 mM NaCl, 5% (v/v) glycerol and 250 mM imidazole was followed by a buffer exchange to phosphate buffered saline using a PD10 column (GE Healthcare). Removal of the His<sub>6</sub> tag with 20 μg l<sup>-1</sup> TEV protease required from 16 h to 5 days at room temperature depending on the construct. TEV protease, His<sub>6</sub> tag and uncleaved protein were removed via a second passage of the sample through Ni<sup>2+</sup> affinity chromatography resin. The eluate was concentrated to 2.5 ml using either an Amicon Ultra-15 (Millipore) or Vivaspin 20 (Sartorius) centrifugal filter unit. Following a final buffer exchange to 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 0.1% sodium azide and 1 mM EDTA with a PD10 column the samples were concentrated to at least 0.2 mM protein. GST-tagged protein was purified using glutathione-agarose chromatography. The sample was bound to the column in 50 mM Tris (pH 8.0), 150 mM NaCl, 2 mM dithiothreitol and 1 mM EDTA, with elution using the same buffer containing 10 mM freshly reduced glutathione. RNA oligonucleotides were purchased from Biospring GmbH.

**Spin labelling.** Residues on the surface of each RRM and distant from the RNA-binding area (namely N155, A164, A171, L187, A188, T209, D273, S281, A287 and A318) were mutated individually to cysteine. The corresponding single cysteine mutant proteins were expressed and purified as described above. Before addition of 3 molar equivalents of 3-(2-iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy radical (iodoacetamido-PROXYL; Sigma-Aldrich) dissolved in methanol, the protein samples were completely reduced by the addition of 2 mM dithiothreitol, and extensively dialysed in 50 mM Tris (pH 8.0) and 200 mM NaCl. Following an overnight reaction in the dark at 4 °C, the modified protein was passed three times through a PD10 desalting column (GE Healthcare Life Sciences) to remove all unreacted spin label and change the buffer to 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 0.1% sodium azide and 1 mM EDTA.

**NMR spectroscopy.** All samples contained 0.2 to 0.8 mM protein in 20 mM sodium phosphate (pH 6.5), 50 mM NaCl with 10% <sup>2</sup>H<sub>2</sub>O added for the lock. Spectra were recorded at 295 K using DRX500, DRX600, AV800 or AV900 Bruker NMR spectrometers, equipped with cryogenic triple resonance gradient probes. Spectra were processed using NMRPipe/Draw<sup>27</sup> and analysed using Sparky 3 (T. D. Goddard and D. G. Kneller, University of California) and NMRView<sup>28</sup>. Protein backbone assignments were obtained from HNCACB and HNCA spectra, or by comparison to related <sup>1</sup>H, <sup>15</sup>N-HSQC and -TROSY spectra and previously published data<sup>11</sup>. Amino acid side chain resonance assignments were obtained from standard HCCH-TOCSY, <sup>15</sup>N- and <sup>13</sup>C-edited NOESY-HSQC experiments. About 15 intermolecular NOEs between the U9 RNA and U2AF65(RRM1–RRM2) were identified for well-resolved peaks in the 3D <sup>13</sup>C-edited NOESY-HSQC experiments. For 13 of these peaks, chemical shifts of the corresponding protein signals could be assigned.

Amide <sup>15</sup>N relaxation data were acquired at 600 MHz and 295 K as described<sup>29</sup>. Steady-state heteronuclear {<sup>1</sup>H}<sup>15</sup>N-NOE spectra were recorded with and without 3 s of <sup>1</sup>H saturation. Relaxation rates and error calculations were determined using NMRView v.4 (ref. 28).

<sup>1</sup>H-<sup>15</sup>N residual dipolar couplings were measured using an interleaved spin-state-selective <sup>1</sup>H, <sup>15</sup>N-TROSY experiment. <sup>15</sup>N-<sup>13</sup>C residual dipolar couplings were measured using a 3D-HNCO experiment<sup>30</sup>. Alignment media consisted of Pf1 phage (Profos AG) or a liquid crystalline mixture of hexanol and pentaethylene glycol monododecyl ether<sup>24</sup>.

Paramagnetic relaxation enhancements (PREs) arising from the spin label were determined using a ratio of peak intensities in the paramagnetic and diamagnetic state (*I*<sub>para</sub>/*I*<sub>dia</sub>) from <sup>1</sup>H, <sup>15</sup>N-HSQC and/or -TROSY spectra without and with the addition of 6 molar equivalents of ascorbic acid. RNA-bound samples included the addition of 1.5 molar equivalents of 10 mM RNA dissolved in water (BioSpring GmbH)<sup>13</sup>. In the case of the N155C mutant, the PRE was also determined directly through the measurement of <sup>1</sup>H *T*<sub>1</sub> and *T*<sub>2</sub> relaxation times<sup>13,31</sup>. A spin label was also incorporated onto an oligoribonucleotide consisting of a 5' 4-thiouridine followed by eight standard uridine residues (BioSpring GmbH), using the same reagent and strategy as detailed above<sup>32</sup>.

**Structure calculation.** Structures were calculated using modified CNS protocols in the ARIA/CNS setup<sup>13,33,34</sup>. In brief, the protocol consists of the following steps: (1) local refinement of the available domain structures of RRM1 and RRM2 using RDC data measured from two alignment media; (2) generation of linker and spin labels, randomization of the linker residues in the RRM1-linker-RRM2 sequence; (3) molecular dynamics simulated annealing restraining RRM1 and RRM2 harmonically to their refined starting structures, with additional dihedral angle restraints from secondary chemical shifts using TALOS<sup>35</sup>, RDCs (omitted for free U2AF65) and hydrogen bond restraints.

Major changes to the standard structure calculation set-up include the generation of the template structures and the randomization protocol: the template structure is generated by reading in available domain crystal structures of RRM1 and RRM2 (ref. 12), which are then fixed by a harmonic energy potential during the simulated annealing protocol. Randomization is restricted to linker residues connecting the two RRM domains, which are kept rigid. The spin label groups are attached to cysteine residues using a patch, which allows the incorporation of one or several (non-interacting) copies of the proxyl moiety to each site. Calculations are performed with an ensemble of four spin labels per cysteine. Simulated annealing protocols and temperature course are the same as in standard structure calculations. The resulting structure ensembles were further refined by replacing the spin-labelled cysteines with the corresponding wild-type residues followed by energy minimization and final refinement in a shell of water molecules<sup>13</sup>.

For residual dipolar couplings, the structures of both RRM domains are refined individually with an effective energy constant for the positional restraints (10 kcal mol<sup>-1</sup> Å<sup>-2</sup>) allowing for local refinement of the protein backbone by the RDC restraints. This step allows slight rearrangements for the backbone atoms of some residues and improves the overall agreement with the RDC data. During the structure calculation of the tandem domain complex, these locally refined structures are then restrained with a very high effective energy constant (non-crystallographic force constant 10,000 kcal mol<sup>-1</sup> Å<sup>-2</sup>). Note that the relative domain orientation of RRM1 and RRM2 does not change if locally refined or unrefined structures are used in the protocol.

Measured intensity ratio from HSQC spectra of oxidized and reduced spin-labelled proteins were converted into paramagnetic relaxation rates and distances as described<sup>13</sup>.

Quality factors for RDC and PRE restraints are calculated as

$$Q = \sqrt{\frac{\sum (V_{\text{backcalc}} - V_{\text{exp}})^2}{\sum (V_{\text{exp}})^2}}$$

where *V*<sub>backcalc</sub> and *V*<sub>exp</sub> are the back calculated and experimental RDC of PRE values for a given structure.

The structural statistics for the structure ensembles with the spin-label molecules still attached are provided in Supplementary Table 2. The structural statistics for the final water-refined closed and the open, U8-bound conformations are given in Supplementary Table 1. The final ensemble of ten structures of the RNA-bound open conformation has 94.6% and 5.2% in the most favoured and additional allowed regions, respectively, for residues within RRM1 (150–229) and RRM2 (260–336). For the ensemble of ten structures of the closed conformation, the same ranges display 93.3% and 6.7% of residues in the most favoured and additional allowed regions, respectively.

**RNA titrations by NMR.** For each RNA titration, samples initially contained 0.2 mM protein in 500 μl of 20 mM sodium phosphate (pH 6.5), 50 mM NaCl, 1% sodium azide and 1 μM EDTA. Chemical shift perturbation was followed by measuring <sup>1</sup>H, <sup>15</sup>N-HSQC and/or -TROSY-HSQC spectra with cumulative addition of 10 mM RNA (BioSpring GmbH) dissolved in H<sub>2</sub>O. Typical titration series used steps of 0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 1.2 and 1.5 molar equivalents of RNA to protein.

**Isothermal titration calorimetry.** ITC was carried out using VP-ITC or ITC200 Microcal calorimeters (Microcal) at 25 °C. All proteins were dialysed extensively using Slide-A-Lyzer 3.5-kDa molecular weight cutoff cassettes (Pierce Biotechnology) against 20 mM sodium phosphate (pH 6.5), 50 mM NaCl and 0.1 mM EDTA. Buffer from the dialysis was used to resolubilize the RNA (BioSpring GmbH).

and to provide a baseline as required. The data were analysed using program Origin version 5.0 provided by Microcal.

**Surface plasmon resonance.** An IAsys resonant mirror biosensor (AffinitySensors) was used to determine the equilibrium constants for the interactions of biotinylated U9 oligonucleotide with RRM1, RRM2 and RRM1–RRM2 proteins<sup>36</sup>. The cuvette was prepared by an initial capture of neutravidin on the biotin-coated surface, and subsequent attachment of the biotinylated U9 RNA. Nonspecifically bound U9 was removed by washing with 2 M NaCl followed by washes with PBS, containing 0.1% Tween20 and binding buffer (20 mM sodium phosphate at pH 6.5, 50 mM NaCl and 1  $\mu$ M EDTA). Following cuvette equilibration with binding buffer, association phase binding responses were recorded separately for various protein concentrations with subsequent washing of the cuvette and monitoring of the dissociation phase. The sensor surface was regenerated by sequential washing with 2 M NaCl and binding buffer. As a negative control for binding experiments immobilized neutravidin was used. All experiments were performed at 20 °C. The experimental data were corrected for nonspecific binding and analysed by using the FASTfit software provided by the manufacturer.

**Singular value decomposition analysis.** To determine the fraction of the unbound and bound populations of RRM1–RRM2 with the different Py tract RNA, the PRE data were fitted as a linear combination of the PRE data of the free RRM1–RRM2, and those of the U9-bound RRM1–RRM2. Only residues for which the PRE data were used as restraints in the calculation of the free and U9-bound structures, and which show inter-domain bleaching, were considered. For the spin label attached to residue 318, 51 residues in RRM1 were included in the fit, with the results expressed as the fraction of U9-bound conformation. Error analysis consisted of 100 iterations of a Monte Carlo simulation with error added only to the experimental data and not the two models.

**In vitro splicing assays.** Pre-spliceosome A complex assembly was carried out as described previously<sup>25</sup>. *In vitro* transcribed RNAs corresponded to the 3' half of intron 1 (including the 3'-splice-site region) and part of exon 2 of AdML promoter transcripts. The sequence of the wild-type (U8) is gggaagcuugcugcagcguaggcgcauguaguccagggguuccuugaugaugucauacuuaccuguccuuuuuuuuccacagCUCGCGG UUGAGGACAAACUCUUCGCGGUCUUUCCAGUGGGGAUCC; intron and exon nucleotides are indicated in lower- and uppercase letters, respectively, and the underlined sequence was replaced by uuuuuaaa (U4A4), uuuuuaaaauuuu (U4A4U4) or uuuuuaaaaaauuuu (U4A8U4) to generate the different mutant substrates. 40,000 c.p.m. (20 fmols) of each <sup>32</sup>P-UTP body-radiolabelled RNA substrate for various mutants (RNA integrity and amounts verified by denaturing gel electrophoresis) were incubated with varying amounts of HeLa cell nuclear extracts (CILBIOTECH; ATP depleted by incubation 30 min at 30 °C) supplemented with 3 mM MgCl<sub>2</sub>, 24.9 mM KCl, 3.33% PVA, 13.3 mM HEPES pH 8, 0.13 mM EDTA, 13.3% glycerol, 0.03% NP-40, 0.66 mM DTT and supplemented or not with 2 mM

ATP and 22 mM creatine phosphate in a final volume of 9  $\mu$ l. The mixture was incubated for 5 min at 30 °C (Supplementary Fig. 11a) or for different time points (Supplementary Fig. 11b). 1  $\mu$ l of heparin (10  $\mu$ g  $\mu$ l<sup>-1</sup>) was added and incubated for 10 min at room temperature. 3  $\mu$ l of 50% glycerol were added and 10  $\mu$ l loaded on a composite gel (4% acrylamide, 0.05% bis-acrylamide, 0.5% agarose, 50 mM Tris, 50 mM glycine). The gel was run for 6 h at 200 V in a cold room in 50 mM Tris, 50 mM glycine buffer. The gel was dried and exposed overnight with a PhosphorImager screen. Quantification for experiments using 3  $\mu$ l HeLa cell nuclear extracts with 5 min incubation at 30 °C was carried out using Image Quant v5.2. Complex A formation was tested in nuclear extracts depleted of U2AF by chromatography in oligo-dT cellulose<sup>26</sup> complemented with 2.5, 7 and 22 ng  $\mu$ l<sup>-1</sup> of recombinant purified GST–U2AF65 (WT) or mutant D215R/G319R (Mut). Results were reproducibly obtained with different depleted extracts and recombinant proteins harbouring different tags. Under optimal conditions of depletion and complementation, the wild-type protein was more than 2.5 times more active than the mutant protein.

27. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
28. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
29. Farrow, N. A. *et al.* Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).
30. Yang, D., Venters, R. A., Mueller, G. A., Choy, W. Y. & Kay, L. E. TROSY-based HNC0 pulse sequences for the measurement of 1HN–15N, 15N–13CO, 1HN–13CO, 13CO–13C $\alpha$  and 1HN–13C $\alpha$  dipolar couplings in 15N, 13C, 2H-labelled proteins. *J. Biomol. NMR* **14**, 333–343 (1999).
31. Iwahara, J., Schwieters, C. D. & Clore, G. M. Ensemble approach for NMR structure refinement against <sup>1</sup>H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126**, 5879–5896 (2004).
32. Ramos, A. & Varani, G. A new method to detect long-range protein–RNA contacts: NMR detection of electron–proton relaxation induced by nitroxide spin-labeled RNA. *J. Am. Chem. Soc.* **120**, 10992–10993 (1998).
33. Nilges, M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660 (1995).
34. Brünger, A. T. *et al.* Crystallography 1 NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
35. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
36. Buckle, P. E. *et al.* The resonant mirror: a novel optical sensor for direct sensing of biomolecular interactions part II: applications. *Biosens. Bioelectron.* **8**, 355–363 (1993).



## ERRATUM

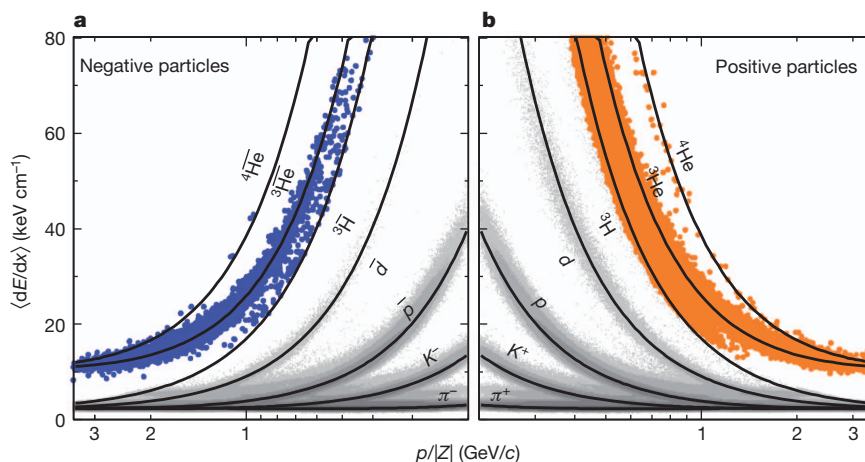
doi:10.1038/nature10264

## Observation of the antimatter helium-4 nucleus

The STAR Collaboration

*Nature* **473**, 353–356 (2011)

In Fig. 2 of this Letter, the lower part of the figure was printed wrongly (the corrected Fig. 2 appears below). The online HTML and PDF versions are correct.



# CAREERS

**TURNING POINT** Bioorganic chemist aims to pay his success forward **p.415**

**INNOVATION** University support boosts productivity for academic inventors **p.415**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



The Mayo Clinic in Rochester, Minnesota, is a powerhouse of biomedical research.

MINNESOTA

## Medicine and materials

*Minnesota made its mark in medical devices and has green potential. But state funding woes could hamper progress.*

BY PAUL SMAGLIK

In 1949, Earl Bakken co-founded a medical-equipment repair company out of his garage in Minneapolis, Minnesota. Nearby, C. Walton Lillehei was pioneering cardiac surgery at the University of Minnesota. The physician, who successfully performed the world's first open-heart surgery in 1952, recognized

that about 1% of children are born with holes in their hearts, and he wanted to find a way to close those holes. To keep the patient's blood flowing while the heart recovered from surgery, Lillehei first tried circulating the blood between the child and their parent, then switched to a plug-in pacemaker. After a power cut in October 1957, Lillehei grew concerned about losing patients during outages.

So he asked Bakken to devise a battery-operated system. Bakken's company grew rapidly, first selling the external pacemaker that he came up with, then developing and marketing an implantable version, as well as defibrillators and other equipment. Medtronic is now based in Fridley, Minnesota, and has 43,000 employees worldwide — with about 10,000 in Minneapolis–St Paul alone.

Entrepreneurial activity in Minnesota has enabled more than 100 medical-device companies to emerge since the middle of last century — many of them direct spin-offs from Medtronic. The industry's growth has been fuelled by intellectual property and research expertise from the University of Minnesota and the Mayo Clinic in Rochester (see 'A promising partnership'). Those successes have sparked burgeoning biomedical and green industries in the region, helping job growth. But a paucity of state funding could be an obstacle — a long-planned major science park has been put on hold, and state support for entrepreneurial ventures is waning.

### TRIGGERING DEVICES

In terms of job growth, "the big horse pulling the sleigh is the device industry", says Dale Wahlstrom, chief executive of the BioBusiness Alliance of Minnesota in St Louis Park. But many of the new jobs springing up in Minnesota are at small companies dealing in biologics and biopharmaceuticals, animal health, food, renewable energy and renewable materials. This signals an increase in opportunities: Wahlstrom says that the number of PhD-level life-sciences jobs in the state has grown from 28,889 in 1997 to a projected 35,459 now.

Challenges loom, however. In 2008, the state legislature approved the University of Minnesota's Biomedical Discovery District, a US\$292-million, 65,000-square-metre cluster of research facilities in Minneapolis. The last stage of construction began this May. But the chances of the university receiving funding for a 130,000-square-metre science park that it proposed five years ago are uncertain. Neither the university nor the state government has committed the money necessary for the project to qualify for a federal loan — the deadline for which is coming up. The state government shut down on 1 July, when Governor Mark Dayton and the state legislature failed to agree on fixes to an estimated \$5-billion state budget deficit; the hold-up further jeopardizes the science park. (As *Nature* went to press, a budget resolution was imminent.) ►

► Minnesota's entrepreneurial environment is also something of a dichotomy. A report by the Kauffman Foundation in Kansas City, Missouri (*The 2010 State New Economy Index*, Kauffman Foundation, 2011) ranked the state 7th in the country for industrial investment in research, but only 39th for non-industry investment, including federal and state funding, and 42nd for entrepreneurial activity.

Still, the region shows potential in the medical-device sector and life-sciences ventures. The Kauffman report ranked Minnesota eighth for the percentage of scientists and engineers in the workforce, eighth for the number of advanced degrees among workers and sixth in terms of the percentage of jobs that were managerial, professional and technical positions requiring at least two years of university education.

### CROSSING OVER

Wahlstrom notes that the state's life-sciences specialities are increasingly diversifying and expanding. Companies that make biomaterials using clean chemistry are starting to have a role in developing medical devices. 3M, a materials company based in Maplewood, Minnesota, has a medical-device division, and Cargill, a food company based in Minneapolis, helped to pioneer green chemistry by creating Natureworks,

a biopolymer manufacturer based in Minnetonka, in 1997.

The Mayo Clinic and the University of Minnesota both feed talent into companies big and small, and some Minnesota entrepreneurs who made their fortunes with legacy companies are providing funds for start-up firms. Steve Oesterle, senior vice-president for medicine and technology at Medtronic, says that Minnesota has a "deep well" of such angel investors, who are encouraged by a state tax credit introduced last year.

Entrepreneur Manny Villafana is one successful businessman who is giving back. He created seven medical-device companies after leaving Medtronic in 1971 and influenced others. For example, Vascular Solutions in Minneapolis is a spin-off from another of his companies, ATS Medical, which Medtronic purchased last year. Vascular Solutions develops, manufactures and markets devices that aid in peripheral heart surgery. The company has about 350 employees, many of them biomedical engineers and material scientists, and has been growing by about 20 people a year.

Minnesota's science jobs aren't all in the medical-device field. Talent, money and intellectual property are also flowing into green chemistry from the state's larger companies and academic

institutions. Chemical engineers, molecular biologists and materials scientists are building firms that aim to reduce waste and pollution in fuels, materials and chemical processing. Some of these companies have attracted venture capital despite the slow economic recovery, allowing them to recruit scientists.

For example, BioAmber of Plymouth, Minnesota — a Cargill spin-out company — aims to turn agricultural crops into chemicals that can replace petroleum-based plastics. It raised \$45 million this year, and is looking to build a production plant that will employ fermentation engineers, molecular biologists and analytical chemists. Jeff Warwick, BioAmber's director of analytical chemistry, says that the company plans to hire ten scientists this year. BioAmber intends to draw from the chemistry talent pools at Cargill, 3M and the University of Minnesota. "We will aggregate more and more resources in terms of people, so it's important that we have local connections," says Warwick. Another Minnesota biomaterials company, Segetis in Golden Valley, has 25 employees, and last month raised the money to expand to as many 100 when its production plant opens later this year, says a company spokeswoman.

The state seems poised to emerge as a player in green technology, says Doug Cameron, founder and managing director of Alberti Advisors, a firm based in Plymouth that matches venture capital to companies. He points to BioAmber's venture-capital success, as well as to increasing sales at Segetis and Natureworks and the purchase of a biofuels plant in Luverne, Minnesota, by Gevo of Englewood, Colorado.

## DIABETES DIVIDENDS

### *A promising partnership*

Two of Minnesota's biggest research institutions are hoping that past success in collaboration will bode well for future ambitions. In 2004, the state started funding the Minnesota Partnership for Biotechnology and Medical Genomics, hoping that a joint venture between traditional rivals the University of Minnesota in Minneapolis and the Mayo Clinic in Rochester would attract federal money. So far, it has worked, with a state investment of more than US\$90 million drawing over \$100 million in federal grants.

Now the institutions want to run a similar project, but with a focus on diabetes. They aim to raise even more money, in what administrators call a moon shot — a bid to make substantial advances in treating, preventing and curing diabetes over the next 10 years. It could also mean substantial opportunities for diabetes researchers.

Launched last October, the state's 'Decade of Discovery' project aims to garner up to \$200 million in money from the state government, private companies and philanthropic organizations, and an equivalent amount from the National Institutes of Health and other US government funders. That will provide hundreds of jobs in diabetes research,

treatment and prevention. Victor Montori, Mayo's director of the Decade of Discovery, says that the diabetes effort was intentionally designed as a job-creation programme, adding that investment — and jobs — will come not just in biomedical research, but also in health insurance, clinics, community organizations and the business community.

Tim Mulcahy, vice-president for research at the University of Minnesota, says that the project is well positioned to attract money. It will benefit from a wealth of basic research at the university, clinical expertise at Mayo, input from the medical-device community, and interactions with public-health and public-policy experts. But even so, Eric Wieben, a biochemist at Mayo, admits that fund-raising will be a challenge, given the state's — and the nation's — current economic condition. "We are working with a two-year election cycle," he says. "We need to show that there is a return on this."

Wieben says that the two institutions can point to the earlier successes of the Minnesota Partnership to show that the diabetes programme will work. Mayo has added hundreds of research positions since the partnership's inception, despite the national recession. **P.S.**

### CHALLENGES AFOOT

But Minnesota may find that a legacy of medical devices and a few successful ventures don't necessarily translate into overall success for sciences start-ups. According to Tim Mulcahy, vice-president of research at the University of Minnesota, the state needs to take steps to nurture innovation and entrepreneurship — and attract new jobs. This includes helping early-stage companies find federal support, helping entrepreneurs network with investors, and connecting academia with industry to establish centres of research excellence. These recommendations were part of a report (*Minnesota Science and Technology Authority Strategic Plan: Turning Ideas into Jobs* Minnesota Science and Technology Authority, 2011) that Mulcahy and others drafted. Legislators responded favourably at first, says Mulcahy, but state budget battles have pushed such initiatives aside. Although the state has invested in innovation, it does not have a cohesive, coordinated plan, he says. Probable budget cuts will make addressing the elements of an innovation framework more difficult. "All of these things," says Mulcahy, "need to be in place in order to end up with jobs." ■

**Paul Smaglik** is a science writer based in Milwaukee, Wisconsin.



# TURNING POINT

## Christian Hackenberger

M. WIMMER

*Christian Hackenberger received Germany's prestigious Heinz Maier-Leibnitz prize in March for his efforts in finding a way to link proteins to synthetic molecules — an important step in adapting proteins for medical applications. Hackenberger, a bioorganic chemist at the Free University of Berlin, discusses how incremental successes gave him the confidence to pursue big projects.*

### When did you decide you wanted to be a research chemist?

When I was a child, my parents worked in the chemical industry. They didn't tell me to study chemistry, but their careers influenced me. At school, testing ideas in a laboratory fascinated me. I decided to study at the University of Freiburg in Germany.

### You also studied in the United States. What did you gain from that experience?

It was a shortcut. I went to the United States for the first time in 1998, with a one-year scholarship to study organic chemistry at the University of Wisconsin–Madison. This was after the first part of my German undergraduate degree. I had my own project, crystallizing the proteins in membranes and using them to make new detergents. I earned a master's degree, although I had gone simply hoping to do some research. With that degree, I could skip the German diploma study — the second part of the degree system before Germany adopted bachelor's and master's degrees. And I could start on my PhD straight away. After my PhD, I did a postdoc at the Massachusetts Institute of Technology (MIT) in Cambridge.

### How did you get involved in science communication?

I moved to Aachen University in Germany to do my PhD, using chemistry techniques to synthesize proteins with specific three-dimensional shapes. My supervisor, Carsten Bolm, gave me a lot of freedom. He let me spend three months as an intern at Westdeutscher Rundfunk, a television station in Cologne, where I wrote for science shows. Ultimately, I decided to focus on research, but my communications experience was not wasted. Especially in chemistry, you need people who can report their research in a way that other people can understand.

### What was the most pivotal moment in the course of your research?

At MIT I had to find a way to change the structures of proteins using a mixture of



techniques. There was a big problem with making one particular protein — the technique just didn't work. I was under a lot of pressure because my scholarship was about to run out, and the group that we collaborated with was waiting for some samples. I had to dive into the literature to come up with new ways to manipulate the proteins, and I managed to devise a technique at the last minute. Because I had succeeded this time, I knew that I could succeed in the future.

### What led to your getting the Heinz Maier-Leibnitz prize?

I received more recognition after my group discovered a chemical reaction that allows peptides and proteins to be manipulated and altered. During the past three years I've given more than 40 lectures and talks in Europe, the United States and Canada. I was also the first person to receive a grant for untenured academic researchers from the Boehringer Ingelheim Foundation in Heidesheim. That helped me to get the professorship at my university. Then I received the Heinz Maier-Leibnitz prize, which is really prestigious because it is not awarded solely for chemistry. Prizes are very important for advancement in the German system.

### Do you plan to stay in Germany?

I have been well funded here, and received a lot from this community. I would be very happy to give something back by helping to support young scientists. I have organized a national collaborative network in chemical biology — young investigators are unlikely to be able to work both chemically and biologically at the beginning of their careers, so we hope to bring individuals from different backgrounds together to work on joint projects. ■

INTERVIEW BY KATHARINE SANDERSON

## DEVELOPING WORLD

### International funding

The US National Science Foundation (NSF) and the Agency for International Development have opened a funding stream for scientists in the developing world. The Partnerships for Enhanced Engagement in Research (PEER) will enable collaborations with scientists who are funded by the NSF; the US National Academies will help to administer the initiative. Applicants need a letter of support from their US-based partners. The first request for proposals will be released in August, and the first round of funding will be awarded later this year. Six PEER pilot projects — focused on areas such as hydrology, biodiversity and seismology — are already being financed in Tanzania, Bangladesh and elsewhere.

## INNOVATION

### Support breeds patents

A supportive atmosphere helps university-based innovators to produce more patents and inventions, says a survey (E. M. Hunter *et al. Res. Pol.* <http://dx.doi.org/10.1016/j.respol.2011.05.024>; 2011) of scientists at Engineering Research Centers (ERCs) — interdisciplinary centres funded by the US National Science Foundation (NSF) to bridge academia and industry. Support includes rewards for commercialization that are built into the tenure or promotion processes, institutional leadership that fosters cross-disciplinary opportunities and technology-transfer offices that streamline the patent process. Universities trying to move away from publications as the sole metric of promotion could use ERCs as a model, says study author Emily Hunter, an organizational psychologist at Baylor University in Waco, Texas.

## UNITED STATES

### University outlooks

US universities have mixed economic outlooks, according to a survey (see [go.nature.com/2ttkoc](http://go.nature.com/2ttkoc)) of 480 private and public colleges by *The Chronicle of Higher Education* and Moody's Investors Service, market analysts in New York. About 90% of public universities faced declines in financial support, compared with just 21% of private universities. But only 1% of private and 5% of four-year public institutions were very likely to enforce mandatory unpaid leave in 2011–12; and 6% of private and public four-year colleges were very likely to freeze faculty hiring.

# EVENT HORIZON

*A bitter pill.*

BY JEFF HECHT

The sad eyes of the well-dressed, neatly trimmed man looked oddly familiar, but it was his question that gave him away. “Found any little green men lately?”

Alexa couldn’t believe it until she glanced at his convention badge. “Karl! It’s been ages!” He had been scrawny and scruffy when they did astrobiology postdocs together a quarter-century ago. “What are you doing now?” She had lost track of him after she landed a tenure-track job.

“I gave up searching,” he said, the sadness in his voice matching that in his eyes.

“What happened to your models of how advanced civilizations would develop?”

“I never published anything,” he said. “After the fellowship fell through, I got a job building computer models of economic trends. It pays the bills. I talked about long-term market models here at the Futures conference.”

“Oh?” Alexa hadn’t noticed his name.

“It was in the business sessions. I wouldn’t expect you to notice. I saw you were in the far-future session, so I decided to come. I owe you something.”

Alexa looked at him blankly.

“I didn’t think you would remember. We bet a meal on who would get a job first. When you did, I was too broke to take you to McDonald’s. I can afford a nice dinner now, and I want to hear what you’re doing.”

“That would be great,” Alexa said. With no travel budget and another big jump in her medical insurance, she had been about to skip dinner.

“What’s surprising is that so many factors in the Drake equation are so favourable for extraterrestrial intelligence,” Alexa said, enjoying the restaurant’s ambience. “Remember how the first hot Jupiters were so exciting when we were postdocs? Now we’ve got tens of thousands of terrestrial planets in habitable zones. There has to be life out there.” She paused to sip the wine, a vintage that she never could have afforded.

“But where are the

little green men?” Karl asked.

“They’re out there, but we haven’t found them yet. No signals at radio or optical frequencies. Maybe they’re sending something we can’t detect, neutrinos or particles we’ll never know about until they turn the colliders back on.”

“What are the odds on intelligence?”

“Once you get multicellular animals, models show intelligence is likely in a few



hundred million years. Technology just needs hands. Dolphins and elephants are out of luck, but bears could do it. Maybe squirrels would have a chance.”

“But how long can technological civilizations last? We haven’t blown ourselves up yet, but our resources are going up in smoke, and we have fouled the nest badly. How much of your annual carbon allowance did you burn getting here?”

“We don’t need to visit other worlds. Sending laser or radio-frequency signals uses hardly any energy compared with interstellar travel. Anybody within 90 light years can listen to old broadcasts of *The Lone Ranger*.”

“But that isn’t free. Who’ll pay the bills when the foundations go broke?” Karl set down his wine glass. “My models tell me where the money is going. Do you know what’s the fastest-growing part of the economy?”

“Climate controls and energy?” Alexa guessed.

Karl shook his head. “Health care. It hit 32% of the US gross domestic product last year. It was only 5% in 1960. That’s a sixfold increase in fraction of the GDP in 70 years, and the economy has expanded a factor of 100 over that time, so dollar spending is up a factor of 600.”

Alexa thought of her insurance bill. “But medicine has improved tremendously. We conquered polio and measles, and we can manage diabetes. Life expectancy is longer than ever.”

“Barely,” Karl muttered. “The marginal return on investment has been declining for decades.” He slipped a mobile from his pocket and flashed a chart on the tablecloth with its nano-projector. A line labelled ‘life expectancy’ levelled off, but one marked ‘medical spending’ rose exponentially.

“What does that mean?”

“The economy is approaching a medical event horizon. Better technology makes other products cheaper with time, but medicine has hit fundamental limits. Instead of buying more and more, each health-care dollar buys less and less. Companies advertise instant muscle tone-ups while you watch 3D, and people buy them although they aren’t nearly as good as exercise. Our economy is spiralling into a black hole; all new

resources go into health care. NASA can’t get a penny for the 15-metre space telescope.”

“But that’s just a temporary delay until we can work out the budget deficit.”

“You can dream,” Karl said. “But that’s how any advanced civilization will behave. It’s entirely rational for intelligent beings to try to maintain their own health and extend their own lifetimes. China slashed human space exploration to fund better health care. Little green men will do the same, so they will never land on the White House lawn.”

“But we can still listen for them,” Alexa protested.

“You can try,” Karl sighed. “Maybe one of our neighbours in the Galaxy will broadcast their version of *The Lone Ranger* long enough for us to hear it.” ■

**Jeff Hecht** is Boston correspondent for *New Scientist* and a contributing editor to *Laser Focus World*.

JACEY

**GO NATURE.COM**  
Follow Futures on  
Facebook at:  
[go.nature.com/mtoodm](http://go.nature.com/mtoodm)



# Photoentrainment and pupillary light reflex are mediated by distinct populations of ipRGCs

S.-K. Chen<sup>1</sup>, T. C. Badea<sup>2</sup> & S. Hattar<sup>1,3</sup>

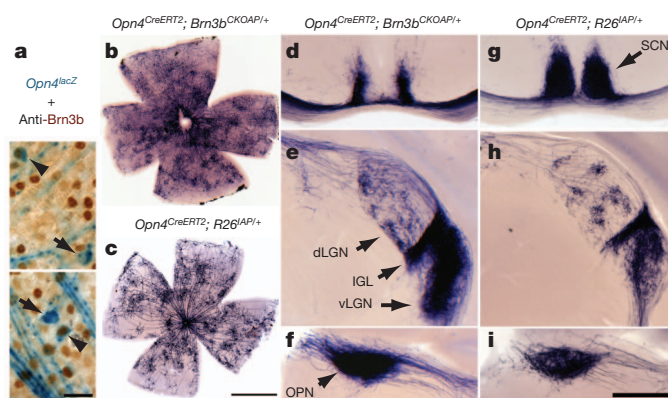
**Intrinsically photosensitive retinal ganglion cells (ipRGCs) express the photopigment melanopsin and regulate a wide array of light-dependent physiological processes<sup>1–11</sup>. Genetic ablation of ipRGCs eliminates circadian photoentrainment and severely disrupts the pupillary light reflex (PLR)<sup>12,13</sup>. Here we show that ipRGCs consist of distinct subpopulations that differentially express the *Brn3b* transcription factor, and can be functionally distinguished. *Brn3b*-negative M1 ipRGCs innervate the suprachiasmatic nucleus (SCN) of the hypothalamus, whereas *Brn3b*-positive ipRGCs innervate all other known brain targets, including the olivary pretectal nucleus. Consistent with these innervation patterns, selective ablation of *Brn3b*-positive ipRGCs severely disrupts the PLR, but does not impair circadian photoentrainment. Thus, we find that molecularly distinct subpopulations of M1 ipRGCs, which are morphologically and electrophysiologically similar, innervate different brain regions to execute specific light-induced functions.**

In addition to rod and cone photoreceptors, the retina contains a small subset of ipRGCs that express the photopigment melanopsin<sup>1,9</sup>. ipRGCs project to the suprachiasmatic nucleus (SCN) and the olivary pretectal nucleus (OPN), regions in the brain that control circadian rhythms and the pupillary light reflex (PLR), respectively. In the absence of the melanopsin protein (*Opn4*), ipRGCs lose their intrinsic photosensitivity<sup>7</sup>, but still innervate the correct brain regions<sup>7</sup> and convey rod/cone input<sup>7,14,15</sup> to drive non-image-forming visual functions<sup>7,16</sup>. Recent studies have shown that ipRGCs are not uniform and can be further subdivided into distinct subtypes based on their morphology, electrophysiology and discrete brain targets<sup>2,17</sup>. M1 ipRGCs can be readily distinguished from other ipRGC subtypes (herein referred to as non-M1 ipRGCs) because they are the only subtype with exclusive dendritic stratification in the OFF sublamina of the inner plexiform layer (IPL) in the retina<sup>18,19</sup>. The prevailing view is that M1 ipRGCs are a homogeneous population that send collateral axonal branches to two relay nuclei, the SCN and OPN, to drive circadian photoentrainment and PLR<sup>20</sup>. Genetic ablation of ipRGCs by diphtheria toxin (in *Opn4*<sup>ΔDTA</sup> mice) eliminates circadian photoentrainment and disrupts PLR<sup>12</sup>. Here we surprisingly found that M1 ipRGCs are not a uniform population, but consist of functionally distinct subpopulations defined by their expression of the POU domain transcription factor, *Brn3b*.

Previously, we showed that *Brn3b* (also known as *Pou4f2*) mutant mice, which lack 80% of RGCs, have pronounced deficits in PLR, but are still capable of weak photoentrainment<sup>21</sup>. These findings raise the possibility that the remaining *Brn3b*-negative M1 ipRGCs selectively mediate photoentrainment. To determine the extent of *Brn3b* expression in the M1 ipRGC population, we performed anti-*Brn3b* immunohistochemistry on retinas from *Opn4*<sup>lacZ</sup> mice, together with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) staining that labels only M1 ipRGCs<sup>21</sup>. A fraction of β-galactosidase-positive RGCs stained for *Brn3b* in the adult retina (Fig. 1a).

To determine the projections of *Brn3b*-positive ipRGCs, we mated mice in which an inducible Cre recombinase is driven by the melanopsin

promoter (*Opn4*<sup>CreERT2/+</sup>) to mice having either a ubiquitous Cre-dependent alkaline phosphatase (AP) reporter (*Rosa26-IAP*)<sup>22</sup> or a conditional *Brn3b* knock-in (*Brn3b*<sup>CKOAP/+</sup>)<sup>21</sup> in which Cre recombination causes the alkaline phosphatase coding region to be expressed by the *Brn3b* promoter (Supplementary Fig. 1). Tamoxifen injections in *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> animals result in labelling of M1 and non-M1 ipRGCs by alkaline phosphatase histochemistry (Fig. 1c), but only of *Brn3b*-expressing ipRGCs in *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> animals (Fig. 1b and Supplementary Table 1). Alkaline phosphatase labelling of *Brn3b*-positive ipRGCs allowed us to analyse the dendritic arborizations and central projections of these cells, independently of *Brn3b*-negative ipRGCs (Fig. 1b, d–f). Many *Brn3b*-positive ipRGCs had dendrites arborizing in the ON sublamina of the IPL similar to previous observations for non-M1 ipRGCs<sup>2,18,20,23</sup>. This indicates that *Brn3b* expression is not just restricted to M1 ipRGCs, but is also expressed in non-M1 ipRGCs (Fig. 1b). Comparing the labelling of *Brn3b*-positive M1 and non-M1 ipRGCs with all ipRGC subtypes, we find that most brain targets of ipRGCs show similar patterns of innervation (Fig. 1d–i). In particular, the OPN is innervated fully in both cases (Fig. 1f, i). However, a notable difference is found in the SCN; in the *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> mice, the SCN was completely innervated by alkaline-phosphatase-positive ipRGC fibres (Fig. 1g) similarly to previous studies<sup>2</sup>. In contrast, in the *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> mice,



**Figure 1 | Co-expression of melanopsin and *Brn3b* defines a specific set of ipRGCs.** **a**, Retinal flat mounts from *Opn4*<sup>lacZ</sup> mice stained with anti-*Brn3b* antibody (brown) and X-gal staining (blue) show *Brn3b*-positive (arrowheads; 140 *Brn3b*-positive ipRGCs from 988 lacZ<sup>+</sup> cells, *n* = 5) and *Brn3b*-negative (arrows), M1 ipRGCs. **b–i**, Alkaline phosphatase histochemistry of retina (**b** and **c**) and coronal brain sections (**d–i**) from *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> mice (**b**, **d–f**), or from *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> mice (**c**, **g–i**). Suprachiasmatic region shows partial innervation in *Opn4*<sup>CreERT2/+</sup>; *Brn3b*<sup>CKOAP/+</sup> mice (**d**), compared to full innervation of the SCN in *Opn4*<sup>CreERT2/+</sup>; *R26*<sup>IAP/+</sup> mice (**g**). Both mouse lines show significant ipRGC projections to the IGL and vLGN, and sparse innervation to the dLGN (**e** and **h**) and intense labelling of the OPN (**f** and **i**). Scale bars are 25 μm (**a**), 1 mm (**b** and **c**), and 400 μm (**d–i**).

<sup>1</sup>Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA. <sup>2</sup>Retinal Circuit Development & Genetics Unit, N-NRL/NEI/NIH, Bethesda, Maryland 20892, USA. <sup>3</sup>Department of Neuroscience, Johns Hopkins University-School of Medicine, Baltimore, Maryland 21218, USA.

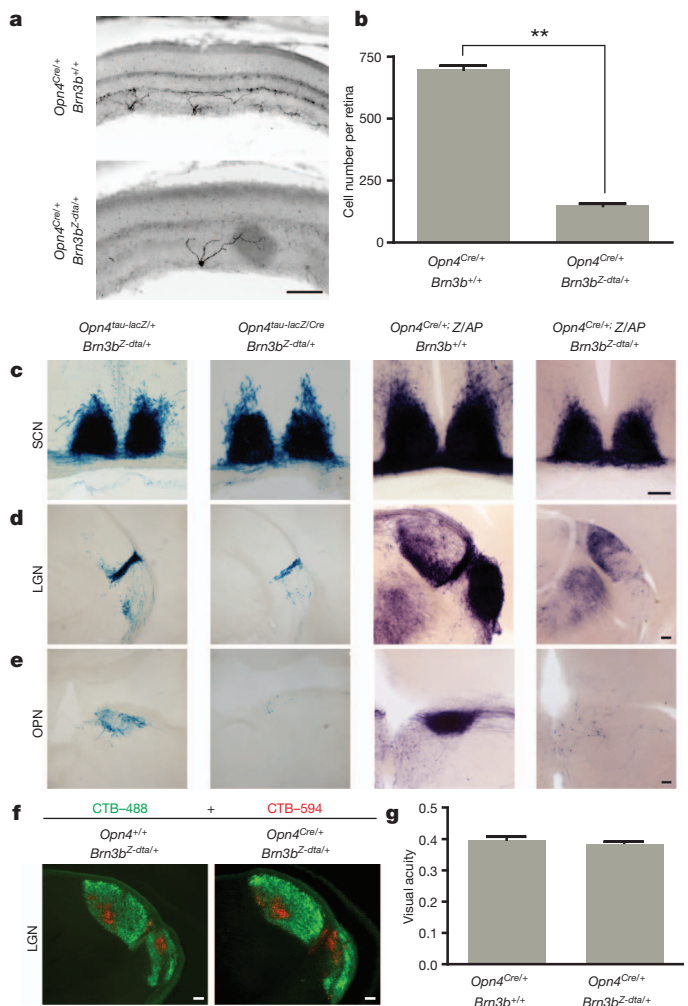


the SCN was sparsely innervated by Brn3b-positive ipRGCs (Fig. 1d), with the medial regions of the SCN completely devoid of innervation (Fig. 1d). We further confirmed that the diminished SCN innervation is not due to the use of the inducible *Opn4<sup>CreERT2</sup>* line, because crossing the *Opn4<sup>Cre</sup>* line<sup>2</sup> with the *Brn3b* knock-in allele (*Opn4<sup>Cre</sup>; Brn3b<sup>CKOAP/+</sup>* animals) also results in reduced SCN innervation (Supplementary Fig. 2). Thus, ipRGCs can be separated into two subpopulations based on their Brn3b expression and connectivity.

To label the central projection of Brn3b-negative ipRGCs and determine the physiological functions of both Brn3b-negative and Brn3b-positive ipRGCs, we specifically eliminated cells that co-express melanopsin and Brn3b by crossing *Opn4<sup>Cre</sup>* and *Brn3b<sup>Z-dta</sup>* lines (Supplementary Table 1). The *Brn3b<sup>Z-dta</sup>* knock-in line<sup>24</sup> expresses diphtheria toxin A subunit (DTA) from the endogenous *Brn3b* gene promoter (Supplementary Fig. 1) only in the presence of Cre<sup>25</sup>. Thus, in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice, Brn3b-expressing ipRGCs are ablated, whereas Brn3b-negative ipRGCs and conventional (melanopsin negative) RGCs are left intact (Supplementary Fig. 3). Using melanopsin immunofluorescence that only reveals M1 ipRGCs in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* retinas, we observed less than 200 surviving M1 ipRGCs (Fig. 2a, b). To determine the extent of ablation of all ipRGCs in the *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice, we generated triple heterozygous *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>; Z/AP* mice (Supplementary Table 1). Alkaline phosphatase labelling in the *Opn4<sup>Cre/+</sup>; Z/AP* mice reveals all M1 and non-M1 ipRGCs (~2,000 cells)<sup>2</sup>. Using alkaline phosphatase histochemistry in the triple heterozygous mice, we observed similar numbers of surviving ipRGCs as with melanopsin immunofluorescence (Fig. 2a). These results show that all non-M1 cells are ablated and that the surviving 10% (200 out of 2,000) of total ipRGCs are Brn3b-negative and predominantly belong to the M1 subtype.

To assess the central projections of these surviving M1 Brn3b-negative ipRGCs, we crossed *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice with the either the *Opn4<sup>tau-LacZ</sup>* reporter<sup>19,20</sup> or *Z/AP* reporter<sup>26</sup>. Although only 200 M1 ipRGCs remained in the *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* or *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>; Z/AP* mice, we observed that their fibres completely innervated the SCN at levels comparable to those observed in the control groups (Fig. 2c). However, innervation of the intergeniculate leaflet (IGL) was highly attenuated (Fig. 2d) and OPN projections were completely abolished (Fig. 2e). Interestingly, the shell of the OPN showed no fibres in the *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* compared to control mice (Fig. 2e). Given that both the SCN and OPN shell are innervated by M1 ipRGCs<sup>5,20</sup>, differential labelling of these ipRGC targets in *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* mice shows that the M1 subtype of ipRGCs is not a uniform population. To ensure that RGCs that are not intrinsically photosensitive are intact in the *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice, we used cholera toxin injection in the eye to label all RGC fibres in the brain anterogradely. RGCs innervated the dorsal and ventral lateral geniculate nuclei (dLGN and vLGN) normally in these mice (Fig. 2f). This is further supported by the similar visual acuity measured in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* and wild-type mice (Fig. 2g).

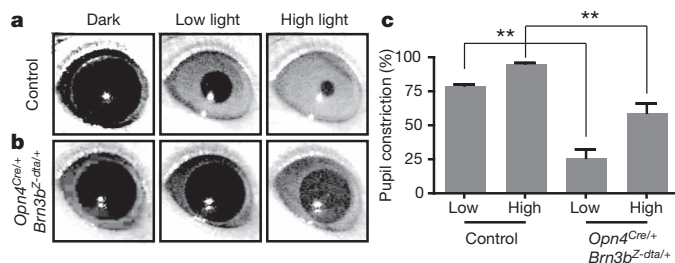
Given that ipRGC projections to the OPN are lost, but SCN projections are largely intact, the *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice allow the relative contributions of Brn3b-negative ipRGCs to the pupillary light reflex (PLR) and circadian light responses to be determined. We first measured PLR in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice at two light intensities in the middle of the day (zeitgeber time 8, ZT8). The pupil of wild-type mice is 95.61% constricted under high light intensity and 79.47% under low light intensity (Fig. 3a, c). In contrast, *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice showed a highly attenuated PLR at ZT8 under high and low light intensities (Fig. 3b, c). This phenotype is remarkably similar to the PLR deficits observed in *Opn4<sup>aDTA/aDTA</sup>* homozygous animals<sup>12</sup>. We further investigated the PLR in the middle of the night (ZT20), and found that *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals have no pupillary constriction to high or low light stimulations (Supplementary Fig. 4 and Supplementary Text). The residual PLR response at ZT8



**Figure 2 | Genetic ablation of Brn3b-positive ipRGCs does not impair targeting to the SCN.** **a**, Melanopsin immunofluorescence reveals a reduction in ipRGC numbers in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* retina compared to control (*Opn4<sup>Cre/+</sup>; Brn3b<sup>+/+</sup>*). **b**, Quantification of surviving melanopsin-positive cells in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* ( $149.8 \pm 8.65$  cells per retina;  $n = 4$ ) and control ( $698.8 \pm 16.85$  cells per retina;  $n = 4$ ) mice. **c–e**, Coronal brain sections of *Opn4<sup>tau-LacZ/+</sup>; Brn3b<sup>Z-dta/+</sup>* and *Opn4<sup>Cre/tau-LacZ</sup>; Brn3b<sup>Z-dta/+</sup>* (**c–e**, left two panels), and *Opn4<sup>Cre/+</sup>; Brn3b<sup>+/+</sup>; Z/AP* and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>; Z/AP* (**c–e**, right two panels) mice using X-gal (left two panels) or alkaline phosphatase histochemistry (**c–e**, right two panels). Sections show SCN (**c**), LGN (**d**) and OPN (**e**). **f**, Labelling of all RGCs with Alexa Fluor 594- and 488-conjugated cholera toxin B (CTB-488 and CTB-594, respectively) in the left (red) and right eye (green), respectively, shows normal brain targeting to image forming regions. **g**, Visual acuity was the same between *Opn4<sup>+/+</sup>; Brn3b<sup>Z-dta/+</sup>* ( $n = 5$ ) and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice ( $n = 6$ ). Scale bars are 100  $\mu$ m. Error bars represent s.e.m.

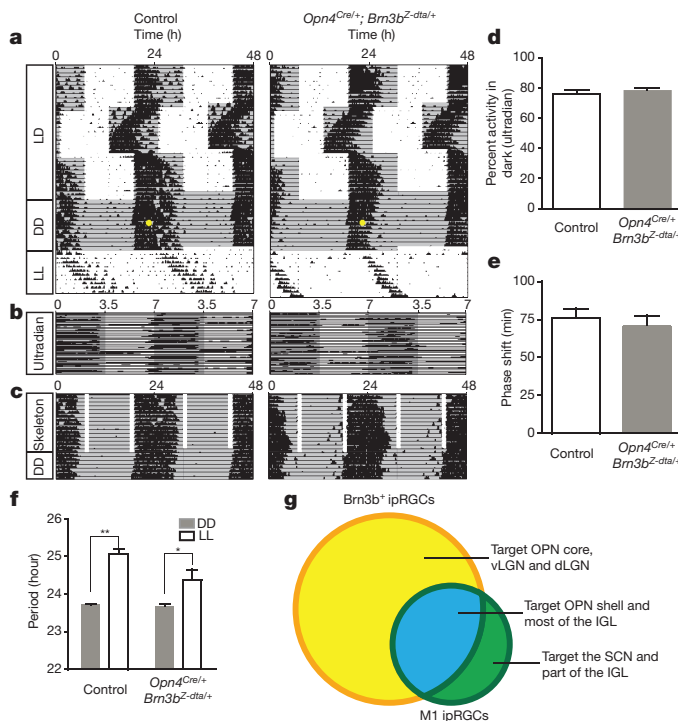
in *Opn4<sup>aDTA/aDTA</sup>* and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice suggests that other melanopsin-negative RGCs contribute to this reflex. One candidate population could be Brn3a-positive RGCs, which project to the OPN<sup>21</sup>.

We then asked whether the surviving M1 ipRGCs that innervate the SCN in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice are sufficient to drive circadian photoentrainment. Strikingly, we found that *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice are able to photoentrain as well as controls under normal 24-h light dark cycles or skeleton photoperiods (Fig. 4a, c, Supplementary Figs 5 and 6 and Supplementary Text). In addition, they can readjust to a 'jet lag' light-dark cycle paradigm with advanced and delayed dark onsets (Fig. 4a). We also observed no difference in activity during the dark phase of the ultradian<sup>16</sup> 3.5:3.5 light/dark (LD) cycle (Fig. 4b, d). Moreover, a 15-min light pulse presented early during the active phase



**Figure 3 | *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice show severe deficits in the pupillary light reflex (PLR).** **a, b,** Representative images of PLR from control and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice. Left panels show pupils under dark conditions, middle panels show pupils under low light intensity ( $22 \mu\text{W cm}^{-2}$ ) and right panels show pupils under high light intensity ( $5.66 \text{ mW cm}^{-2}$ ). **c,** Quantification of PLR data from control ( $n = 5$ ) and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* ( $n = 6$ ) animals. \*\* indicates  $P < 0.01$  with one-way ANOVA. Error bars represent s.e.m.

of mice maintained under constant conditions generated similar phase shifts ( $76.2 \pm 6.5$  and  $70.67 \pm 7.3$  min for control and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals, respectively; Fig. 4e). Together, these results indicate that Brn3b-negative ipRGCs, comprising only 10% of all identified ipRGC subtypes, are sufficient for circadian photoentrainment.



**Figure 4 | *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice show normal circadian photoentrainment.** **a–c,** Representative actograms from control and *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals under a series of lighting paradigms: **a,** 12:12 h LD cycle, 'jet lag', constant darkness (DD), and constant light (LL); **b,** ultradian 3.5:3.5 h cycles; **c,** skeleton photoperiod. The grey background indicates darkness and the yellow dot indicates the 15-min light pulse at CT16 (circadian time 16). *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* animals have similar photoentrainment to controls with minor deficits in period lengthening. **d,** Percent activity in the dark portion of the ultradian cycle shows no significant difference between the genotypes. **e,** Quantification of phase shifts shows no significant differences between the two groups. **f,** Quantification of circadian period from the two groups under constant dark and constant light conditions. Both groups of animals show significant period lengthening under constant light. **g,** Venn diagram showing Brn3b-positive ipRGCs in yellow and Brn3b-negative ipRGCs in green (full description is provided in Supplementary Table 1). \*\* indicates  $P < 0.01$ , \* indicates  $P < 0.05$  using Student's *t*-test. Error bars represent s.e.m.

However, *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice show a minor deficit in period lengthening under constant light conditions (Fig. 4f). Because period lengthening is positively correlated with light intensity, attenuated projections to the IGL in *Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>* mice could underlie this difference.

Here we show that, although M1 ipRGCs have homogeneous morphological and electrophysiological characteristics, they consist of at least two different subpopulations, which can be discriminated by expression of the Brn3b transcription factor (Fig. 4g). The two M1 subpopulations have distinct brain targets and are involved in different non-image forming visual functions. Using precise molecular genetic tools to ablate Brn3b-expressing ipRGCs, we disrupted the pupillary light reflex but not circadian photoentrainment. Thus, ipRGCs have parallel pathways for controlling non-image forming functions, analogous to the specialized properties of RGCs that mediate image-forming functions<sup>27</sup>.

## METHODS SUMMARY

**Animals.** All experiments were conducted in accordance with NIH guidelines and approved institutional animal care and use committees of the Johns Hopkins University.

**Behavioural analyses.** We used previously described behavioural tests<sup>12</sup> that measure visual acuity (optomotor), pupil constriction (PLR), the period of the circadian oscillator (wheel running activity), the adjustment of the circadian clock to different light stimulations (circadian photoentrainment, 'jet lag' paradigms, phase shifting, and skeleton photoperiod) and direct light effects on activity (constant light and ultradian).

**Histology.** X-gal and alkaline phosphatase histochemistry were performed as described previously<sup>2,5,28</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 March; accepted 18 May 2011.

Published online 17 July 2011.

- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Ecker, J. L. et al. Melanopsin-expressing retinal ganglion-cell photoreceptors: cellular diversity and role in pattern vision. *Neuron* **67**, 49–60 (2010).
- Gooley, J. J., Lu, J., Fischer, D. & Saper, C. B. A broad role for melanopsin in nonvisual photoreception. *J. Neurosci.* **23**, 7093–7106 (2003).
- Hannibal, J. & Fahrenkrug, J. Melanopsin containing retinal ganglion cells are light responsive from birth. *Neuroreport* **15**, 2317–2320 (2004).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Hattar, S. et al. Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 75–81 (2003).
- Lucas, R. J. et al. Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
- Panda, S. et al. Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
- Provencio, I., Rollag, M. D. & Castrucci, A. M. Photoreceptive net in the mammalian retina. *Nature* **415**, 493 (2002).
- Ruby, N. F. et al. Role of melanopsin in circadian responses to light. *Science* **298**, 2211–2213 (2002).
- Tu, D. C. et al. Physiologic diversity and development of intrinsically photosensitive retinal ganglion cells. *Neuron* **48**, 987–999 (2005).
- Güler, A. D. et al. Melanopsin cells are the principal conduits for rod-cone input to non-image-forming vision. *Nature* **453**, 102–105 (2008).
- Hatori, M. et al. Inducible ablation of melanopsin-expressing retinal ganglion cells reveals their central role in non-image forming visual responses. *PLoS ONE* **3**, e2451 (2008).
- Schmidt, T. M., Taniguchi, K. & Kofuji, P. Intrinsic and extrinsic light responses in melanopsin-expressing ganglion cells during mouse development. *J. Neurophysiol.* **100**, 371–384 (2008).
- Wong, K. Y., Dunn, F. A., Graham, D. M. & Berson, D. M. Synaptic influences on rat ganglion-cell photoreceptors. *J. Physiol. (Lond.)* **582**, 279–296 (2007).
- Mrosovsky, N. & Hattar, S. Impaired masking responses to light in melanopsin-knockout mice. *Chronobiol. Int.* **20**, 989–999 (2003).
- Brown, T. M. et al. Melanopsin contributions to irradiance coding in the thalamo-cortical visual system. *PLoS Biol.* **8**, e1000558 (2010).
- Berson, D. M., Castrucci, A. M. & Provencio, I. Morphology and mosaics of melanopsin-expressing retinal ganglion cell types in mice. *J. Comp. Neurol.* **518**, 2405–2422 (2010).
- Hattar, S. et al. Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *J. Comp. Neurol.* **497**, 326–349 (2006).

20. Baver, S. B., Pickard, G. E. & Sollars, P. J. Two types of melanopsin retinal ganglion cell differentially innervate the hypothalamic suprachiasmatic nucleus and the olivary pretectal nucleus. *Eur. J. Neurosci.* **27**, 1763–1770 (2008).
21. Badea, T. C., Cahill, H., Ecker, J., Hattar, S. & Nathans, J. Distinct roles of transcription factors Brn3a and Brn3b in controlling the development, morphology, and function of retinal ganglion cells. *Neuron* **61**, 852–864 (2009).
22. Badea, T. C. *et al.* New mouse lines for the analysis of neuronal morphology using CreER(T)/loxP-directed sparse labeling. *PLoS ONE* **4**, e7859 (2009).
23. Schmidt, T. M. & Kofuji, P. Functional and morphological differences among intrinsically photosensitive retinal ganglion cells. *J. Neurosci.* **29**, 476–482 (2009).
24. Mu, X. *et al.* Ganglion cells are required for normal progenitor- cell proliferation but not cell-fate determination or patterning in the developing mouse retina. *Curr. Biol.* **15**, 525–530 (2005).
25. Gan, L., Wang, S. W., Huang, Z. & Klein, W. H. POU domain factor Brn-3b is essential for retinal ganglion cell differentiation and survival but not for initial cell fate specification. *Dev. Biol.* **210**, 469–480 (1999).
26. Lobe, C. G. *et al.* Z/AP, a double reporter for Cre-mediated recombination. *Dev. Biol.* **208**, 281–292 (1999).
27. Wässle, H. Parallel processing in the mammalian retina. *Nature Rev. Neurosci.* **5**, 747–757 (2004).
28. Badea, T. C., Wang, Y. & Nathans, J. A noninvasive genetic/pharmacologic strategy for visualizing cell morphology and clonal relationships in the mouse. *J. Neurosci.* **23**, 2314–2322 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Nathans for providing several animal lines (*Brm3b<sup>CKOAP</sup>*, *R26<sup>IAP</sup>* and *Z/AP*) that were crucial for the completion of this study. We thank J. L. Ecker, who created the inducible cre line (*Opn4<sup>CreERT2</sup>*) we used in this study. We thank Z. Yang in D. Zack's laboratory for providing the *Brm3b<sup>Z-dta</sup>* mouse line, which was generously provided by the original laboratory that created this line: W. Klein. We also thank R. Kuruvilla, H. Zhao, M. Halpern, A. P. Sampath and T. Schmidt for their careful reading of the manuscript and helpful suggestions and the Johns Hopkins University Mouse Tri-Lab for support. This work was supported by the National Institutes of Health grant GM076430 (S.H.), the David and Lucile Packard Foundation (S.H.), and the Alfred P. Sloan Foundation (S.H.).

**Author Contributions** S.-K.C., T.C.B. and S.H. performed all experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.H. ([shattar@jhu.edu](mailto:shattar@jhu.edu)) or T.C.B. ([tudor.badea@nih.gov](mailto:tudor.badea@nih.gov)).



## METHODS

**Mice.** All mice were of a mixed background (BL/6;129SvJ). Littermate male animals that were used in the behavioural analyses were aged between 4 and 12 months. Animals were housed and treated in accordance with NIH and IACUC guidelines, and used protocols approved by the Johns Hopkins University Animal Care and Use Committees.

**Generation of *Opn4<sup>CreERT2</sup>* line.** To generate *Opn4<sup>CreERT2</sup>* mice, we used the targeting arms and general strategy detailed in ref. 2. The only difference is that the construct contained a rabbit  $\beta$ -globin intron, CreERT2 recombinase and an IRESLacZ cassette immediately downstream of the start codon for mouse melanopsin.

**Immunohistochemistry.** Mouse retina was fixed as whole eyecup for at least 30 min in 4% paraformaldehyde (PFA) and cryoprotected in 30% sucrose overnight. Retina sections (40  $\mu$ m) were obtained by cryostat and incubated with blocking solution (0.3% Triton X-100 and 5% normal goat serum in PBS) for 1 h before staining with primary antibody overnight at 4 °C. Sections were washed in 1  $\times$  PBS three times for 30 min and incubated with secondary antibody at room temperature for 2 h before mounting in vector-shield mounting solution. Images were taken with an Olympus microscope by epifluorescence. The dilution for melanopsin antibody (Advanced Targeting Systems) is 1:1,000.

**Tamoxifen injections.** The intensity of labelling depends on the amount of tamoxifen injected into animals as well as the efficiency of excision from loxP regions in the reporter mice. Therefore, all intraperitoneal injections of tamoxifen were standardized to label all the identified ipRGC subtypes (M1–M5). In fact, ipRGCs with morphologies characteristic for all identified ipRGC subtypes are observed in the flat mount retinas in Fig. 1b, c. Retina shown in Fig. 1b was from an animal injected with 500  $\mu$ g tamoxifen at P14 (postnatal day 14). Brains shown in Fig. 1d–f were from an animal injected with 250  $\mu$ g tamoxifen at P5. For Fig. 1c, g–i, images are from an animal injected with 1 mg tamoxifen at embryonic day 17 (E17). There is no particular reason to injecting tamoxifen at different developmental times. We simply used the alkaline phosphatase staining as a tracing method to reveal ipRGC targets in the brain.

**Histology.** X-gal staining. Mice were perfused with 15 ml of 4% PFA, the brain was dissected out, cryoprotected in 30% sucrose for 2 days and 50  $\mu$ m coronal sections were obtained by cryostat. Brain sections were incubated in staining solution with 1 mg ml<sup>-1</sup> X-gal for 2 days at room temperature, post-fixed in 4% PFA for 1 h and mounted with glycerol<sup>5</sup>.

Alkaline phosphatase (AP) staining. Mice were perfused with 45 ml of 4% PFA, the brain and retina were dissected out. Whole-mount retina was post-fixed for

30 min and 200  $\mu$ m coronal brain sections were obtained by vibrotome. Both retina and brain sections were heat-inactivated at 65 °C for 90 min and incubated in alkaline phosphatase staining solution. After staining, the sections and retina whole-mount were post-fixed in 4% PFA for overnight and washed with ethanol series before mounting<sup>2, 28</sup>.

**Cholera toxin injections in the eye.** Mice were anaesthetized with avertin. Eyes were injected intravitreally with 2  $\mu$ l of cholera toxin B subunit conjugated with Alexa Fluor 488 (CTB-488) or Alexa Fluor 594 (CTB 594) (Invitrogen). Three days after injection, brains were isolated, sectioned and mounted.

**Visual acuity.** A virtual cylinder OptoMotry (Cerebral Mechanics) was used to determine visual acuity by measuring the image-tracking reflex of mice. A sine-wave grating was projected on the screen rotating in a virtual cylinder. The animal was assessed for a tracking response on stimulation for about 5 s. All acuity thresholds were determined by using the staircase method with 100% contrast.

**Pupillary light reflex.** All animals were kept under 12:12 LD cycle before testing PLR. Before each experiment, all animals were dark-adapted for at least 1 h. While one eye received light stimulation with specific intensity described in the main text from a 470-nm light-emitting-diode light source (Super Bright LEDs), a digital camcorder (DCRHC96; Sony) was used to record from the other eye (for 30 s) at 30 frames per second under a 940-nm light (LDP). The percentage pupil constriction was calculated as the percentage of pupil area at 30 s after initiation of the stimulus (steady state) relative to the dilated pupil size (right before light stimulation). The same group of animals were used for wheel-running activity. The control animals are littermates to the experimental animals (*Opn4<sup>Cre/+</sup>; Brn3b<sup>Z-dta/+</sup>*) with either *Opn4<sup>Cre/+</sup>; Brn3b<sup>+/+</sup>* or *Opn4<sup>+/+</sup>; Brn3b<sup>Z-dta/+</sup>* genotypes.

**Wheel-running activity.** Mice were placed in cages with a 4.5-inch running wheel, and their activity was monitored with VitalView software (MiniMitter). The period was calculated with ClockLab (Actimetrics). Mice were initially placed under 12:12 LD cycle for 2 weeks. Animals were then exposed to two “jet lag” light paradigms: 10 days of a 6-h advance followed by 10 days of a 6-h delay. After the “jet lag” paradigms, mice were kept under constant darkness for 2 weeks followed by 10 days of constant light. Phase-shifting experiments were carried out on the 7th day of constant darkness where each animal was exposed to a 15 min light pulse at CT16 (1,500 lx). Animals were re-entrained to 12:12 LD cycle for 2 weeks before exposing them to ultradian 3.5:3.5 light/dark cycles. The intensity of light for all the light dark cycle were ~1,000 lx. Another set of mice was tested using a skeleton photoperiod, where two 1-h light pulses (800 lx) separated by 10 h of dark were administered.

# The crystal structure of GXGD membrane protease FlaK

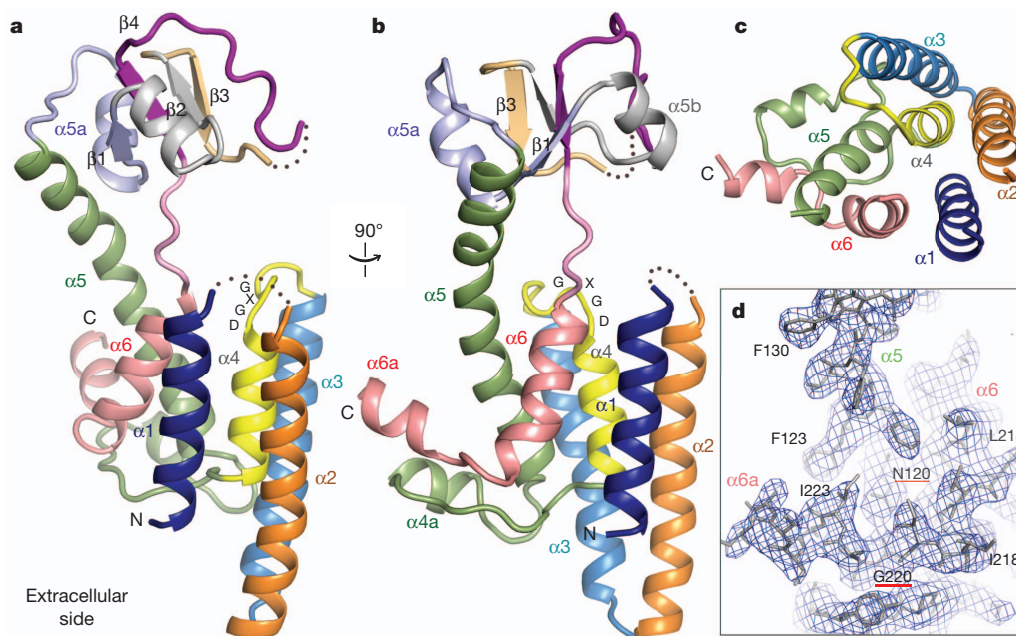
Jian Hu<sup>1</sup>, Yi Xue<sup>1</sup>, Sangwon Lee<sup>1</sup> & Ya Ha<sup>1</sup>

The GXGD proteases are polytopic membrane proteins with catalytic activities against membrane-spanning substrates that require a pair of aspartyl residues<sup>1–4</sup>. Representative members of the family include preflagellin peptidase, type 4 prepilin peptidase, presenilin and signal peptide peptidase. Many GXGD proteases are important in medicine. For example, type 4 prepilin peptidase may contribute to bacterial pathogenesis<sup>5–7</sup>, and mutations in presenilin are associated with Alzheimer's disease<sup>8–10</sup>. As yet, there is no atomic-resolution structure in this protease family. Here we report the crystal structure of FlaK, a preflagellin peptidase from *Methanococcus maripaludis*, solved at 3.6 Å resolution. The structure contains six transmembrane helices. The GXGD motif and a short transmembrane helix, helix 4, are positioned at the centre, surrounded by other transmembrane helices. The crystal structure indicates that the protease must undergo conformational changes to bring the GXGD motif and a second essential aspartyl residue from transmembrane helix 1 into close proximity for catalysis. A comparison of the crystal structure with models of presenilin derived from biochemical analysis reveals three common transmembrane segments that are similarly arranged around the active site. This observation reinforces the idea that the prokaryotic and human proteases are evolutionarily related<sup>11,12</sup>.

The crystal structure presented here provides a framework for understanding the mechanism of the GXGD proteases, and may facilitate the rational design of inhibitors that target specific members of the family.

Archaeal preflagellins and bacterial type-4 prepilins, both of which are type-II (N<sub>in</sub>-C<sub>out</sub>) membrane proteins<sup>13,14</sup>, are synthesized with short, positively charged leader peptides<sup>5,15</sup>. They are cleaved in the membrane, at a site a few residues upstream of the hydrophobic membrane-spanning sequence, by preflagellin peptidase (PFP) and type-4 prepilin peptidase (TFPP), respectively, before being secreted and incorporated into the mature flagellum or type-4 pilus (Supplementary Fig. 1a). We have crystallized FlaK, a prototypic archaeal PFP from *M. maripaludis*<sup>16</sup>. The membrane protease maintained a native-like conformation throughout crystallization because both the crystallization drop and the dissolved crystals showed robust enzymatic activity (Supplementary Fig. 1b). The structure was determined by single-wavelength anomalous dispersion using a Se-Met-substituted crystal (Supplementary Fig. 2 and Supplementary Table 1).

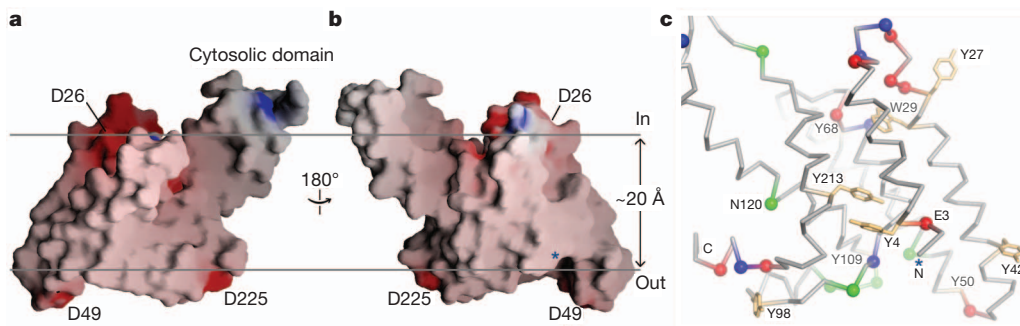
The crystal structure shows that FlaK contains two compactly folded domains: a mostly  $\alpha$ -helical membrane-spanning domain and a soluble domain with four anti-parallel  $\beta$ -strands (Fig. 1; the soluble domain is disordered in one of the two FlaK molecules in the asymmetric unit).



**Figure 1 | The structure of FlaK.** Shown here are cartoon representations of molecule A, one of the two FlaK molecules in the asymmetric unit. In molecule B, the soluble domain cannot be completely traced owing to disorder. These illustrations, as well as those in Figs 2c, 3a, b and 4b, c, were generated by PyMOL (<http://www.pymol.org>). a, b, Two views of the molecule from the side.

The secondary structural elements and the GXGD motif are labelled. c, A view from the cytosolic side of the membrane. For clarity, the soluble domain (including part of  $\alpha 5$ ) is removed. d, A portion of the final  $2F_o - F_c$  electron density map (contoured at 1 $\sigma$  level) showing the position of  $\alpha 6a$  relative to TM helix  $\alpha 5$ .

<sup>1</sup>Department of Pharmacology, Yale School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA.



**Figure 2 | FlaK is tilted in the membrane.** Molecule B is shown in this figure because its TM region is better defined in the electron density map, and does not contain any breaks. **a, b**, The molecular surface, colour-coded by electrostatic potential. The two horizontal lines roughly mark the hydrophobic belt around the membrane protease. The N terminus of the protein is labelled

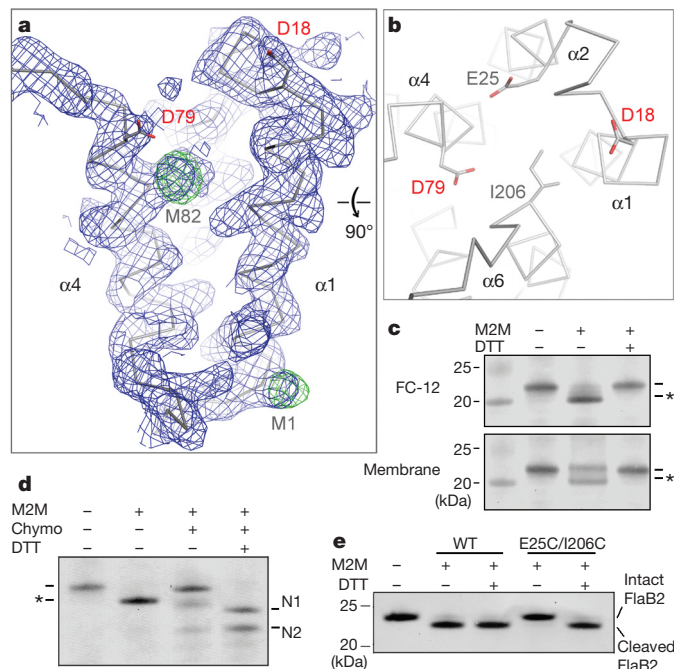
Previous predictions correctly located all six transmembrane (TM) segments<sup>4</sup>, but the crystal structure is more complex (Supplementary Fig. 3a). One major deviation from the prediction occurs around TM4 (yellow in Fig. 1). The hydrophilic loop between TM3 and TM4 does not protrude into the cytoplasm as predicted; instead, it lowers towards the centre of the bundle of TM helices, and is followed by a short TM helix,  $\alpha 4$ . The last TM helix ( $\alpha 6$ ) is also short (pink in Fig. 1), and seems unable to cross the lipid bilayer completely. The protein segments immediately after  $\alpha 4$  and  $\alpha 6$  form an unusual structure that protrudes sideways from the base of the TM helices. This feature does not seem to be an artefact of crystallization. A comparison between the two copies of FlaK in the asymmetric unit shows that the unusual structure is identically positioned, despite the fact that it is involved in different crystal packing interactions (Supplementary Fig. 3b). The amphipathic nature of the structure is also consistent with its position next to the TM helices, and with the possibility that it may interact peripherally with the membrane. In the big loop between  $\alpha 4$  and  $\alpha 5$  (including  $\alpha 4a$ ), and in the carboxy-terminal segment (including  $\alpha 6a$ ), all the polar side chains point downwards away from the membrane, whereas most hydrophobic side chains point up or sideways to interact either with the TM helices, or with lipids that surround the helices. Furthermore, there is a conserved asparagine on  $\alpha 5$  (Asn 120), the side chain of which points outwards to form a hydrogen bond with the carbonyl oxygen of Gly 220 from the extended segment between  $\alpha 6$  and  $\alpha 6a$  (Fig. 1d). If the carboxy-terminus of the protein were positioned elsewhere, Asn 120 would become unfavourably exposed to the lipid.

To accommodate the unusual peripheral structure ( $\alpha 4a$  and  $\alpha 6a$ ), and the short TM helix  $\alpha 6$ , the other TM helices must be tilted in the membrane. The tilting is required to avoid positioning charged groups such as Asp 26, Asp 49 and Asp 225 in the hydrophobic region of the membrane (Fig. 2). The tilting also makes  $\alpha 6$  roughly perpendicular to the membrane plane, enabling it to go through the lipid bilayer with the shortest distance. A more thorough examination of the distribution of amino acids in the TM region supports the tilted model. As shown in Fig. 2c, the lower boundary of the membrane is roughly marked by a thin belt of acidic residues (red), basic residues (blue), asparagines and glutamines (green). Four tyrosine residues (Tyr 42, Tyr 50, Tyr 98 and Tyr 109) are also found within this belt. Tyrosine and tryptophan often cluster to the interface between water and lipid<sup>17</sup>. The upper boundary of the membrane probably corresponds to a plane that goes through Tyr 27, Trp 29 and Tyr 68. Although Glu 3, Tyr 4, Asn 120 and Tyr 213 appear between the two boundaries, a closer inspection shows that the polar groups on their side chains are not directly exposed to the lipid. According to the tilted model, the membrane has to be constricted around the protease. This feature was previously thought to be unique to the intramembrane proteases<sup>18–20</sup>.

The two aspartyl residues that are essential for catalysis (Asp 18 and Asp 79) are located at the ends of TM helices  $\alpha 1$  and  $\alpha 4$ , respectively.

by a blue asterisk. **c**, The  $C_{\alpha}$  trace of the TM region with the same orientation as in panel **b**. Red spheres, negatively charged residues; blue spheres, positively charged residues; green spheres, asparagines and glutamines. Tyrosines and tryptophans are shown as orange stick models (undefined side chains are not shown).

Despite the relatively low resolution at which the crystal structure was solved,  $\alpha 1$  and  $\alpha 4$  are clearly defined in the electron density map, and the experimentally determined selenium sites in both helices are consistent with the register of protein sequence with the density (Fig. 3a). The positions of  $\alpha 1$  and  $\alpha 4$  are almost identical in the two independently modelled FlaK molecules in the asymmetric unit of the crystal (Supplementary Fig. 3b). The spatial relationship between the two



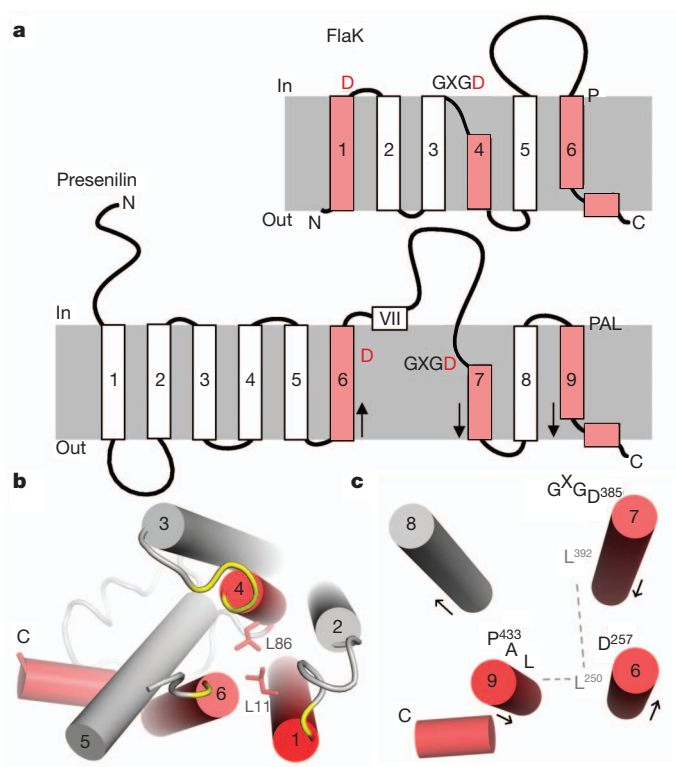
**Figure 3 | The uncoupling between Asp 18 and Asp 79.** **a**, The  $2F_o - F_c$  map (contoured at  $1\sigma$  level; blue) and the anomalous Fourier map (contoured at  $4\sigma$  levels; green) around  $\alpha 1$  and  $\alpha 4$ . **b**, A view from the cytosolic side of the membrane. The distance from  $C_{\alpha 1}$  of Ile 206 to  $C_{\alpha 7}$  of Glu 25 is  $\sim 6$  Å. **c**, Crosslinking causes the E25C/I206C mutant protein to migrate faster in the SDS–polyacrylamide gel (to a position marked by the asterisk); adding DTT breaks the disulphide linkage. FC-12, foscholine-12. The left lane is a molecular weight marker. **d**, Partial proteolysis confirms the covalent linkage between Cys 25 and Cys 206. Chymotrypsin (chymo) cleaves FlaK twice between TM domains 5 and 6, generating two N-terminal fragments (N1 and N2). For crosslinked mutant protein, the N-terminal fragments (containing Cys 25) remain attached to the C-terminal fragment (containing Cys 206). **e**, Crosslinking between Cys 25 and Cys 206 renders the protease completely inactive; treatment with DTT fully restores activity. Wild-type (WT) FlaK is not affected by M2M and DTT. The antibody used in the western blots (panels **c**–**e**) detects the N-terminal His tag of FlaK and the C-terminal His tag of FlaB2 (substrate).



helices is, however, surprising from a functional point of view because it creates a 12 Å gap between Asp 18 and Asp 79. To investigate whether the FlaK crystal structure represents a non-active conformation of the protease, we introduced a pair of cysteines (E25C and I206C) to the two TM helices ( $\alpha 2$  and  $\alpha 6$ ) that are on the opposite side of the gap (Fig. 3b). The double mutant (E25C/I206C) was proteolytically active, indicating that its conformation was not markedly perturbed by the mutations. As shown in Fig. 3c, d, the two cysteines can be readily crosslinked by 1,2-ethanedithiol dimethanesulphonate (M2M). Given its short length, the crosslinker has to lie inside the gap between  $\alpha 1$  and  $\alpha 4$  to bridge Cys 25 and Cys 206. Taken together, these results indicate that, in the absence of substrate, the membrane protease can adopt an inactive conformation in which the two catalytic aspartyl residues are structurally uncoupled. Consistent with this idea, we found that crosslinking between  $\alpha 2$  and  $\alpha 6$ , which prevents movement of Asp 18 and Asp 79 towards each other, completely eliminated protease activity (Fig. 3e). Breaking the crosslinking disulphide bonds with dithiothreitol (DTT) fully restored protease activity. The observation that crosslinking in the membrane fraction is less complete than in detergent solutions indicates that FlaK may assume additional conformations in the lipid bilayer (Fig. 3c). The behaviour of FlaK is similar to that of presenilin in these regards: the human protease also switches between at least two conformations and in one conformation the two catalytic aspartates do not closely oppose each other<sup>21,22</sup>.

Three regions in the FlaK sequence are highly conserved (Supplementary Fig. 4). The first two are centred on Asp 18 and on the GXGD motif (Asp 79). The third region, which corresponds to the sequence around the amino terminus of TM helix  $\alpha 6$ , is shown by the crystal structure to be near the active site as well. The roles of individual residues from the conserved regions in the protease mechanism were probed by mutagenesis (Supplementary Fig. 5). Three main lessons can be learned from this analysis. First, Glu 23, Glu 25 and Asp 26, the only three acidic residues around the active site, do not seem to be essential for the binding of the positively charged leader peptide. Second, among the three glycine residues in the GXGD motif, Gly 76 is the most critical for function. Because Gly 76 is not closely packed against other residues, the large reduction of activity in the G76A mutant probably results from a partial loss of backbone flexibility in this important region. Third, most mutations from Pro 201 to Pro 208 produce a small effect; P208A, however, is different in that it markedly enhances enzymatic activity. Pro 208 is packed against TM helix  $\alpha 5$ , and points away from the active site. If the conformational change in the protease involves movement of TM helices  $\alpha 1$ ,  $\alpha 4$  and  $\alpha 6$  (Fig. 3b), then the altered packing between  $\alpha 6$  and  $\alpha 5$  might have facilitated such a change.

Presenilin has nine TM segments<sup>23,24</sup> (Fig. 4a). The last four segments (TM6–TM9) of presenilin share homology with signal peptide peptidase, and are thought to constitute the core of the protease<sup>25</sup>. Previous cysteine-scanning mutagenesis and crosslinking experiments indicated that the active site of presenilin is housed in an open hydrophilic cavity<sup>21,26</sup>, surrounded by the two TM segments (TM6 and TM7) that carry the catalytic aspartates (Asp 257 and Asp 385), and by the C-terminal TM9, which bears a conserved Pro-Ala-Leu motif<sup>22,27</sup>. The crystal structure of FlaK now shows that the active site of the prokaryotic protease has a similar architecture. As illustrated in Fig. 4a, the TM segments TM1, TM4 and TM6 of FlaK would be equivalent to presenilin's TM6, TM7 and TM9. In FlaK, there is a highly conserved leucine (Leu 11) seven residues upstream of the catalytic Asp 18 on TM1 (Supplementary Fig. 4). Leu 11 is packed against Leu 86 from TM4 (Fig. 4b). Leu 86, which is also conserved, is seven residues downstream of the catalytic Asp 79. The residues in presenilin corresponding to Leu 11 and Leu 86 would be Leu 250 and Leu 392, respectively (seven residues away from the catalytic Asp 257 and Asp 385). Both Leu 250 and Leu 392 are highly conserved<sup>28</sup>. The fact that Asp 257 and Asp 385 are close to each other indicates that Leu 250 and Leu 392, which are on the same side of the helices as the aspartates, may also



**Figure 4 | Structural comparison between FlaK and presenilin-1.**

**a**, Topology diagrams of FlaK and presenilin. The three key TM segments are highlighted in red. The grey boxes represent membrane. Arrows indicate the direction of the helices (N to C termini). **b**, FlaK viewed from the cytosolic side of the membrane. Asp 18, the GXGD motif and the conserved Pro 204 are shown in yellow. The two conserved leucines are shown as stick models. **c**, Packing of the three key TM helices in presenilin. This model is similar to the one proposed in ref. 27, but is the mirror image of another model<sup>22</sup>. Leu 250 from TM6 is hypothesized to mediate packing with TM7 (through Leu 392) and TM9.

interact (Fig. 4c). In FlaK, Leu 11 makes additional contact with Tyr 213 from TM6. In presenilin, Leu 250 also seems to interact with the last TM segment because it can be readily crosslinked to many positions on TM9<sup>27</sup>. Besides the packing of the three key helices, presenilin and FlaK share other features. TM7 of presenilin, like its counterpart in FlaK, also seems to contain two structural elements: an exposed region bearing the GXGD motif and a short, tightly packed hydrophobic helix<sup>26</sup>. The last TM segments in both proteases have conserved proline residues near the N terminus, and are followed by an amphipathic helix that interacts peripherally with the membrane<sup>27</sup>. It is important to note that presenilin lacks the equivalent of FlaK's TM segments 2 and 3. Therefore, other TM helices may join the central three TM segments to complete the active site. The hydrophobic domain VII, which undergoes endo-proteolysis<sup>1</sup>, must also be bound initially at the active site<sup>22</sup>.

The structure of FlaK's active site is fundamentally different from that of pepsin, a classic aspartyl protease<sup>29</sup>, in that it lacks an internal two-fold symmetry and its two catalytic aspartyl residues are not rigidly fixed. Membrane proteases have evolved unique mechanisms to conduct catalysis inside lipid bilayers. For example, rhomboid serine proteases use a surface cap to control access to a preformed and membrane-embedded Ser-His catalytic dyad<sup>30</sup>. The uncoupling between two catalytic aspartyl residues, indicated by the crystal structure described here and by earlier biochemical studies<sup>21,22</sup>, may represent a general mechanism that is widely adopted by the GXGD proteases. Such an uncoupling mechanism could potentially have an important role in regulating catalysis.

## METHODS SUMMARY

MmarC6\_0338 (encoding FlaK) was amplified by PCR from the genomic DNA of *M. maripaludis* strain C6, cloned into pET-28a and expressed in *Escherichia coli* Rosetta 2 (DE3) cells (Novagen). The Se-Met-substituted membrane protein was extracted in foscholine-12 (Anatrace) and purified using metal-affinity and size-exclusion columns. The concentrated protein ( $\sim 10 \text{ mg ml}^{-1}$ ) was extensively dialysed against a buffer containing 20 mM HEPES (pH 7.3), 100 mM NaCl, 0.06% Cymal-6 (Anatrace) and 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP). Single crystals were prepared by the sitting-drop method, in which 1  $\mu\text{l}$  of protein solution was mixed with 1  $\mu\text{l}$  of well solution containing 30% PEG 300, 50 mM glycine (pH 9.5) and 100 mM NaCl. Many crystals were screened at the national synchrotron light source (NSLS, X25 and X29) and at the advanced photon source (APS, 24-ID-C and E) before a final data set at 3.6 Å resolution was collected from a single Se-Met-substituted crystal at the selenium peak wavelength, which was used for both phase determination and refinement. The details of protein purification, crystal structure determination, protease activity assay and crosslinking experiments are described in the Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 21 March; accepted 18 May 2011.**

**Published online 17 July 2011.**

- Wolfe, M. S. *et al.* Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and  $\gamma$ -secretase activity. *Nature* **398**, 513–517 (1999).
- LaPointe, C. F. & Taylor, R. K. The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases. *J. Biol. Chem.* **275**, 1502–1510 (2000).
- Weihsen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B. Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* **296**, 2215–2218 (2002).
- Bardy, S. L. & Jarrell, K. F. Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. *Mol. Microbiol.* **50**, 1339–1347 (2003).
- Lory, S. & Strom, M. S. Structure-function relationship of type-IV prepilin peptidase of *Pseudomonas aeruginosa*—a review. *Gene* **192**, 117–121 (1997).
- Craig, L., Pique, M. E. & Tainer, J. A. Type IV pilus structure and bacterial pathogenicity. *Nature Rev. Microbiol.* **2**, 363–378 (2004).
- Sandkvist, M. Type II secretion and pathogenesis. *Infect. Immun.* **69**, 3523–3535 (2001).
- Selkoe, D. J. & Wolfe, M. S. Presenilin: running with scissors in the membrane. *Cell* **131**, 215–221 (2007).
- Jorissen, E. & De Strooper, B.  $\gamma$ -secretase and the intramembrane proteolysis of Notch. *Curr. Top. Dev. Biol.* **92**, 201–230 (2010).
- Brouwers, N., Sleegers, K. & Van Broeckhoven, C. Molecular genetics of Alzheimer's disease: an update. *Ann. Med.* **40**, 562–583 (2008).
- Steiner, H. *et al.* Glycine 384 is required for presenilin-1 function and is conserved in bacterial polytopic aspartyl proteases. *Nature Cell Biol.* **2**, 848–851 (2000).
- Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **36**, D320–D325 (2008).
- Francetic, O., Buddelmeijer, N., Lewenza, S., Kumamoto, C. A. & Pugsley, A. P. Signal recognition particle-dependent inner membrane targeting of the PulG Pseudopilin component of a type II secretion system. *J. Bacteriol.* **189**, 1783–1793 (2007).
- Bayley, D. P. & Jarrell, K. F. Overexpression of *Methanococcus voltae* flagellin subunits in *Escherichia coli* and *Pseudomonas aeruginosa*: a source of archaeal preflagellin. *J. Bacteriol.* **181**, 4146–4153 (1999).
- Kalmokoff, M. L., Karnauchow, T. M. & Jarrell, K. F. Conserved N-terminal sequences in the flagellins of archaeobacteria. *Biochem. Biophys. Res. Commun.* **167**, 154–160 (1990).
- Bardy, S. L. & Jarrell, K. F. FlaK of the archaeon *Methanococcus maripaludis* possesses preflagellin peptidase activity. *FEMS Microbiol. Lett.* **208**, 53–59 (2002).
- Killian, J. A. & von Heijne, G. How proteins adapt to a membrane–water interface. *Trends Biochem. Sci.* **25**, 429–434 (2000).
- Wang, Y., Maegawa, S., Akiyama, Y. & Ha, Y. The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J. Mol. Biol.* **374**, 1104–1113 (2007).
- Bondar, A. N., del Val, C. & White, S. H. Rhomboid protease dynamics and lipid interactions. *Structure* **17**, 395–405 (2009).
- Ha, Y. Structure and mechanism of intramembrane protease. *Semin. Cell Dev. Biol.* **20**, 240–250 (2009).
- Tolia, A., Chavez-Gutierrez, L. & De Strooper, B. Contribution of presenilin transmembrane domains 6 and 7 to a water-containing cavity in the  $\gamma$ -secretase complex. *J. Biol. Chem.* **281**, 27633–27642 (2006).
- Tolia, A., Horre, K. & De Strooper, B. Transmembrane domain 9 of presenilin determines the dynamic conformation of the catalytic site of  $\gamma$ -secretase. *J. Biol. Chem.* **283**, 19793–19803 (2008).
- Laudon, H. *et al.* A nine-transmembrane domain topology for presenilin 1. *J. Biol. Chem.* **280**, 35352–35360 (2005).
- Spasic, D. *et al.* Presenilin-1 maintains a nine-transmembrane topology throughout the secretory pathway. *J. Biol. Chem.* **281**, 26569–26577 (2006).
- Narayanan, S., Sato, T. & Wolfe, M. S. A C-terminal region of signal peptide peptidase defines a functional domain for intramembrane aspartic protease catalysis. *J. Biol. Chem.* **282**, 20172–20179 (2007).
- Sato, C., Morohashi, Y., Tomita, T. & Iwatsubo, T. Structure of the catalytic pore of  $\gamma$ -secretase probed by the accessibility of substituted cysteines. *J. Neurosci.* **26**, 12081–12088 (2006).
- Sato, C., Takagi, S., Tomita, T. & Iwatsubo, T. The C-terminal PAL motif and transmembrane domain 9 of presenilin 1 are involved in the formation of the catalytic pore of the  $\gamma$ -secretase. *J. Neurosci.* **28**, 6264–6271 (2008).
- Ponting, C. P. *et al.* Identification of a novel family of presenilin homologues. *Hum. Mol. Genet.* **11**, 1037–1044 (2002).
- Davies, D. R. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Chem.* **19**, 189–215 (1990).
- Wang, Y. & Ha, Y. Open-cap conformation of intramembrane protease GlpG. *Proc. Natl Acad. Sci. USA* **104**, 2098–2102 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Héroux, H. Robinson and A. Soares at NSLS, and J. Schuermann at APS NE-CAT for their help during data collection. X-ray diffraction data were measured at beamlines X25 and X29 at NSLS, and at 24-ID-C and 24-ID-E at APS. Financial support was principally from the US Department of Energy and from the National Institutes of Health. This work was supported by a New Scholar Award in Aging from the Ellison Medical Foundation (to Y.H.), a gift from the Neuroscience Education and Research Foundation (to Y.H.) and a pilot grant from Yale's programme in Cellular Neuroscience, Neurodegeneration, and Repair (CNR) (to Y.H.).

**Author Contributions** J.H. and Y.X. purified and characterized FlaK in various detergents. J.H. obtained the high-resolution crystals of FlaK. J.H., Y.X. and Y.H. solved the crystal structure. Y.H., Y.X. and J.H. wrote the paper. Y.X. and S.L. screened many constructs and performed the initial biochemical and functional characterizations.

**Author Information** The atomic coordinates of FlaK and structure factors have been deposited in the Protein Data Bank under accession code 3SOX. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to Y.H. ([ya.ha@yale.edu](mailto:ya.ha@yale.edu)).

## METHODS

**Protein expression and purification.** Many PFPs and TFPPs were screened for crystallization. Most of the genes were amplified by PCR from genomic DNAs purchased from ATCC. The genes for *Pseudomonas aeruginosa* PilD/PilA, *E. coli* BfpP and *Dichelobacter nodosus* FimP were gifts from S. Lory, M. Donnenberg and J. Rood, respectively. MmarC6\_0338 (encoding FlaK) was amplified from the genomic DNA of *M. maripaludis* strain C6, and cloned into pET-28a. To facilitate removal of the N-terminal His tag, a Gly-Ser-Gly-Ser sequence was inserted between the thrombin site and FlaK sequence. Mvol\_1295 (encoding FlaB2) was amplified from the genomic DNA of *Methanococcus voltae* strain A3 and cloned into pET-43b with a C-terminal His tag<sup>16</sup>. FlaK was expressed in *E. coli* Rosetta 2 (DE3) cells (Novagen), grown in Luria broth. FlaB2 was similarly over-expressed in C43(DE3) cells (Lucigen). To generate Se-Met FlaK, cells were cultured at 37 °C in M9 minimum media supplemented with Se-Met, then induced by 1 mM  $\beta$ -D-thiogalactopyranoside (IPTG) at an absorbance at 600 nm ( $A_{600}$ ) of 0.6, and grown at 20 °C for 16 h before collection. Cytoplasmic membranes were prepared by the spheroplast method<sup>31</sup> and suspended in a buffer containing 50 mM sodium phosphate (pH 7.2), 500 mM NaCl, 5 mM  $\beta$ -mercaptoethanol and a cocktail of complete protease inhibitors (tablet without EDTA, Roche Diagnostics). For solubilization, powder of foscholine-12 (Anatrace) was added to the membrane suspension to achieve a final concentration of 1% (w/v). The His-tagged protein was eluted from a TALON metal affinity column (Clontech) in 50 mM sodium phosphate (pH 7.2), 500 mM NaCl, 200 mM imidazole, 5 mM  $\beta$ -mercaptoethanol and 0.1% foscholine-12. After passing through a Sephadex G-25 desalting column, the sample was cleaved by thrombin overnight at 22 °C. Finally, the protein was loaded onto a Superdex-200 column (GE Healthcare) equilibrated with 20 mM HEPES (pH 7.3), 100 mM NaCl, 1 mM TCEP and 0.1% foscholine-12. The peak fraction was pooled, concentrated to 10 mg ml<sup>-1</sup> and dialysed against 20 mM HEPES (pH 7.3), 100 mM NaCl, 1 mM TCEP and 0.06% Cymal-6 (Anatrace) at 4 °C for 8 days. About 3 mg of purified membrane protein could be obtained for crystallization trials from 1 litre of bacterial culture.

**Crystallization and structure determination.** The sitting-drop method was used to prepare Se-Met FlaK crystals: 1  $\mu$ l of protein solution (4 mg ml<sup>-1</sup>; the lower concentration is due to precipitation during dialysis) was mixed with 1  $\mu$ l of well solution containing 30% PEG 300, 50 mM glycine (pH 9.5) and 100 mM NaCl. Needle-shaped crystals usually appeared in 2 days at 22 °C and grew to full size in 1 week. The crystals were dehydrated by equilibrating for 24 h against a well solution containing 40% PEG 300, before direct flash-freezing in liquid nitrogen. Screening and data collection were performed at the national synchrotron light source (X25 and X29) and at the advanced photon source (24-ID-C and E). All diffraction data were processed by HKL2000 (ref. 32). The structure was determined by single-wavelength anomalous dispersion<sup>33</sup> using a highly redundant data set which was generated by merging four data sets collected at four different spots on a single Se-Met crystal at 24-ID-C. The same data set was used in refinement (Supplementary Table 1). The selenium sites and the initial phases were determined by hkl2map<sup>34</sup>. The experimental electron density map confirmed the presence of two FlaK molecules in the asymmetric unit, and clearly showed all the TM helices (Supplementary Fig. 2). The soluble domain in molecule A was visible but could not yet be traced; the soluble domain in molecule B was mostly missing. Averaging the TM regions of the two molecules by dm<sup>35</sup> improved the clarity of the map. Modelling of the polypeptide chains using O<sup>36</sup> was assisted by known Se sites (Supplementary Fig. 6). After rounds of model building and refinement by CNS<sup>37</sup>, the phases were sufficiently improved to allow complete tracing of the soluble domain in molecule A. The final model was refined by CNS and

refmac5 (ref. 38). The electrostatic potential surfaces shown in Fig. 2a, b were generated by GRASP<sup>39</sup>.

**FlaK activity assay.** The enzymatic activity of FlaK was measured according to ref. 4. In brief, membranes containing overexpressed FlaK or FlaB2 were prepared using the spheroplast method, and were re-suspended in phosphate buffer. The membrane fractions were then mixed and the reaction, at 22 °C, was initiated by adding a  $\times 5$  reaction buffer containing 2.5% Triton X-100 and 100 mM HEPES (pH 7.3). The reaction was stopped by mixing with SDS–polyacrylamide gel electrophoresis sample-loading buffer. The reaction mixture was examined by western blot using an anti-His-tag antibody (Calbiochem). The purified FlaK in detergent solutions was assayed similarly. The two assays shown in Supplementary Fig. 1b were conducted at 22 °C for 120 min and 90 min, respectively: in the top panel, a large amount of protease (as indicated by the protease band) and a long incubation time were used to exclude the possibility of residual enzymatic activity in the asparagine mutants. In Supplementary Fig. 5, a shorter assay time (45 min) and a smaller amount of protease were used so that both the intact and processed FlaB2 are visible: in this setting, the amount of intact substrate is a good indicator of the reaction rate. The amount of protease used in the assay was reflected in the lower control panel, where the loading (of FlaK alone) was ten times that used in the assay, to increase the visibility of the protease. The same amount of protease was used in the assay shown in Fig. 3e, but the reaction time was longer (120 min).

**Chemical crosslinking.** The membrane preparation containing the E25C/I206C double mutant was suspended in a buffer containing 50 mM sodium phosphate (pH 8.0) and 100 mM NaCl. Crosslinking was performed either directly in the membrane suspension or in a foscholine-12-solubilized membrane fraction (0.2% foscholine-12), by treating the protein with 2 mM M2M (Toronto Research Chemicals Inc.) at 22 °C for 1 h. M2M has a spacer arm length of 5.2 Å<sup>40,41</sup>. The reaction was stopped by adding an equal volume of 40 mM *N*-ethylmaleimide in 200 mM HEPES (pH 7.3). DTT (final concentration 50 mM) was used to break the crosslinking disulphide bond.

- Kaback, H. R. & Stadtman, E. R. Proline uptake by an isolated cytoplasmic membrane preparation of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **55**, 920–927 (1966).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Wang, B. C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112 (1985).
- Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843–844 (2004).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
- Loo, T. W. & Clarke, D. M. Determining the dimensions of the drug-binding domain of human P-glycoprotein using thiol cross-linking compounds as molecular rulers. *J. Biol. Chem.* **276**, 36877–36880 (2001).
- Green, N. S., Reisler, E. & Houk, K. N. Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. *Protein Sci.* **10**, 1293–1304 (2001).



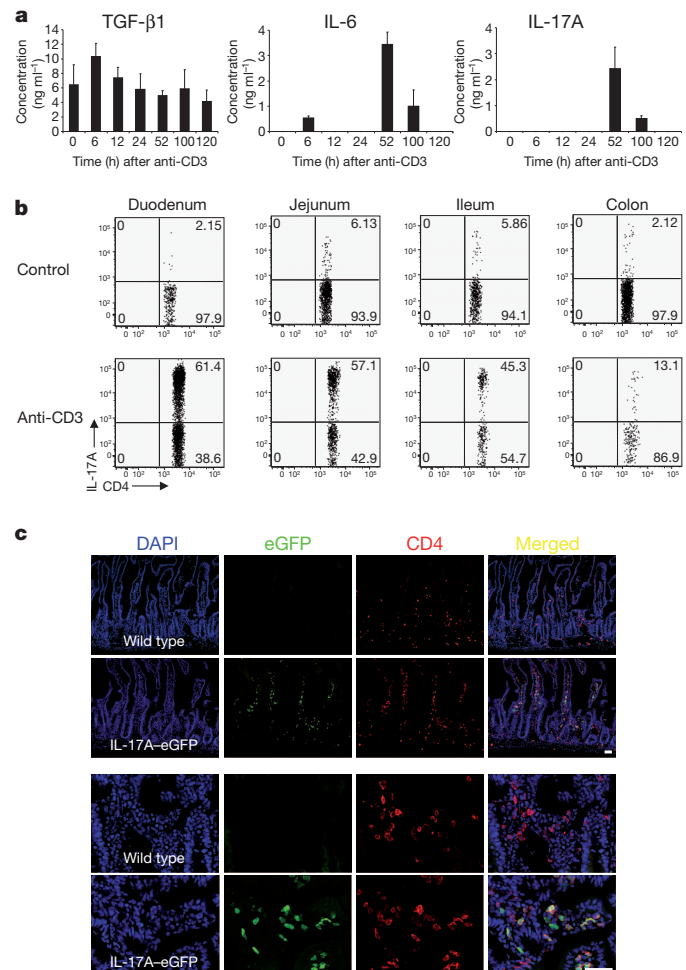
# Control of T<sub>H</sub>17 cells occurs in the small intestine

Enric Esplugues<sup>1,2,3\*</sup>, Samuel Huber<sup>1,4\*</sup>, Nicola Gagliani<sup>5</sup>, Anja E. Hauser<sup>2</sup>, Terrence Town<sup>6,7</sup>, Yisong Y. Wan<sup>8</sup>, William O'Connor Jr<sup>1</sup>, Anthony Rongvaux<sup>1</sup>, Nico Van Rooijen<sup>9</sup>, Ann M. Haberman<sup>10</sup>, Yoichiro Iwakura<sup>11</sup>, Vijay K. Kuchroo<sup>12</sup>, Jay K. Kolls<sup>13</sup>, Jeffrey A. Bluestone<sup>14</sup>, Kevan C. Herold<sup>1</sup> & Richard A. Flavell<sup>1,15</sup>

Interleukin (IL)-17-producing T helper cells (T<sub>H</sub>17) are a recently identified CD4<sup>+</sup> T cell subset distinct from T helper type 1 (T<sub>H</sub>1) and T helper type 2 (T<sub>H</sub>2) cells<sup>1</sup>. T<sub>H</sub>17 cells can drive antigen-specific autoimmune diseases and are considered the main population of pathogenic T cells driving experimental autoimmune encephalomyelitis (EAE)<sup>2</sup>, the mouse model for multiple sclerosis. The factors that are needed for the generation of T<sub>H</sub>17 cells have been well characterized<sup>3–6</sup>. However, where and how the immune system controls T<sub>H</sub>17 cells *in vivo* remains unclear. Here, by using a model of tolerance induced by CD3-specific antibody, a model of sepsis and influenza A viral infection (H1N1), we show that pro-inflammatory T<sub>H</sub>17 cells can be redirected to and controlled in the small intestine. T<sub>H</sub>17-specific IL-17A secretion induced expression of the chemokine CCL20 in the small intestine, facilitating the migration of these cells specifically to the small intestine via the CCR6/CCL20 axis. Moreover, we found that T<sub>H</sub>17 cells are controlled by two different mechanisms in the small intestine: first, they are eliminated via the intestinal lumen; second, pro-inflammatory T<sub>H</sub>17 cells simultaneously acquire a regulatory phenotype with *in vitro* and *in vivo* immune-suppressive properties (rT<sub>H</sub>17). These results identify mechanisms limiting T<sub>H</sub>17 cell pathogenicity and implicate the gastrointestinal tract as a site for control of T<sub>H</sub>17 cells.

T<sub>H</sub>17 cells have been associated with the pathogenesis of several chronic inflammatory disorders, including rheumatoid arthritis and multiple sclerosis<sup>2,7</sup>. To study the cellular and molecular mechanisms that control pathogenicity mediated by T<sub>H</sub>17 cells we first used the CD3-specific antibody treatment model. It is known that CD3-specific antibody treatment induces a 'cytokine storm' and local inflammation mainly in the small intestine<sup>8</sup>. Despite this it has been validated as an *in vivo* model of tolerization<sup>9</sup> and is now under study in human clinical trials<sup>10</sup>. By mimicking antigen, CD3-specific antibody treatment leads to activation-induced cell death (AICD) of T cells<sup>11,12</sup> and consequently a systemic upregulation of IL-6 (ref. 9) and transforming growth factor- $\beta$  (TGF- $\beta$ 1) induced by phagocyte engulfment of apoptotic T cells<sup>13</sup>. In line with these publications, we found that CD3-specific antibody treatment induced an immunoregulatory environment marked by simultaneous expression of TGF- $\beta$ 1 and IL-6 (Fig. 1a). The combination of these cytokines is important for the development of T<sub>H</sub>17 cells *in vitro* and *in vivo* as it has been previously clearly established<sup>3,4</sup>. Accordingly, we found elevated levels of IL-17A in plasma of CD3-specific antibody-treated animals compared to controls (Fig. 1a).

First, we aimed to investigate the source of IL-17A. It has been reported that a few hours after injection of CD3-specific antibody, there is a rapid disappearance of the majority of T cells from the



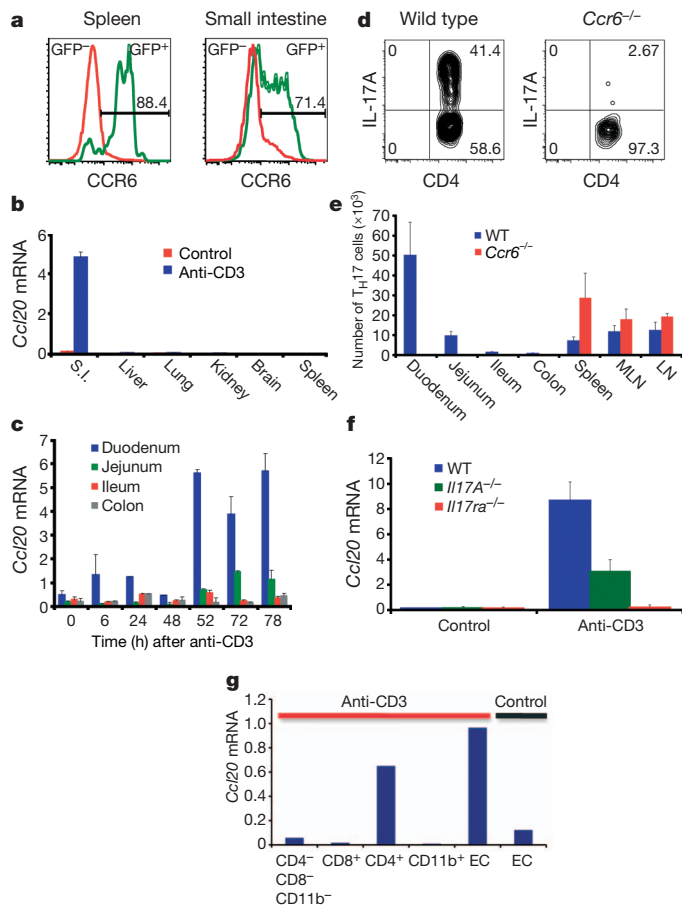
**Figure 1 | Accumulation of T<sub>H</sub>17 cells in the small intestine after CD3-specific antibody treatment.** Mice were injected with CD3-specific antibody. **a**, Plasma levels of TGF- $\beta$ 1, IL-6 and IL-17A. Mean  $\pm$  s.e.m.;  $n = 4$ . **b**, Flow cytometric analysis of IL-17A-eGFP expression (gated on CD4<sup>+</sup>TCR $\beta$ <sup>+</sup> events); numbers in quadrants indicate percent cells in each. **c**, Immunofluorescence staining of frozen sections of the small intestine after CD3-specific antibody treatment (eGFP, green; CD4, red; cell nuclei, DAPI). Scale bar, 50  $\mu$ m. Data are representative of at least three independent experiments.

<sup>1</sup>Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. <sup>2</sup>German Rheumatism Research Center (DRFZ), A Leibniz Institute, Berlin 10117, Germany. <sup>3</sup>Cluster of Excellence NeuroCure, Charité-Universitätsmedizin Berlin, Berlin 10117, Germany. <sup>4</sup>Medizinische Klinik, Universitätsklinikum Hamburg-Eppendorf, Hamburg 20246, Germany. <sup>5</sup>San Raffaele Diabetes Research Institute (HSR-DRI), Milan 20132, Italy. <sup>6</sup>Departments of Biomedical Sciences and Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. <sup>7</sup>Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, California 90048, USA. <sup>8</sup>Lineberger Comprehensive Cancer Center, Department of Microbiology and Immunology, The University of North Carolina, School of Medicine, Chapel Hill, North Carolina 27599, USA. <sup>9</sup>Department of Molecular Cell Biology, Faculty of Medicine, Vrije Universiteit, Amsterdam, 1081 BT, The Netherlands. <sup>10</sup>Department of Laboratory Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA. <sup>11</sup>Center for Experimental Medicine, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. <sup>12</sup>Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>13</sup>Department of Genetics, LSU Health Sciences Center, New Orleans, Louisiana 70112, USA. <sup>14</sup>Diabetes Center at the University of California San Francisco, San Francisco, California 94143, USA. <sup>15</sup>Howard Hughes Medical Institute, New Haven, CT. \*These authors contributed equally to this work.

circulation<sup>13,14</sup>. Surprisingly, in parallel with the disappearance of T cells from the periphery, we found a concomitant increase in the percentage and the number of total T cells in the small intestine, in particular in the duodenum (Supplementary Fig. 1a–c). In a newly generated IL-17A–eGFP knock-in mouse (enhanced green fluorescent protein was inserted in the *Il17a* locus; Methods and Supplementary Fig. 2a–d and 3a–c) injected with CD3-specific antibody, 50–80% of the CD4<sup>+</sup>TCRβ<sup>+</sup> T cells located in the duodenum were expressing IL-17A (Fig. 1b and Supplementary Fig. 1d, e). The percentage and number of T<sub>H</sub>17 cells in the intestine decreased from the duodenum to the colon in a gradient-like fashion (Fig. 1b). Detection of CD4<sup>+</sup>eGFP<sup>+</sup> T cells by immunofluorescence and two-photon-laser-scanning microscopy confirmed the high frequency of T<sub>H</sub>17 cells in the small intestine *in situ* (Fig. 1c and Supplementary Fig. 4a–c). Importantly, we also found T<sub>H</sub>17 cell infiltration in the duodenum when animals were injected with a therapeutic non-FcR-binding CD3-specific antibody<sup>15</sup>, although the frequency and numbers of the T<sub>H</sub>17 cells were lower compared to the FcR-binding antibody (Supplementary Fig. 5a). Similar results were observed after antigen-specific stimulation when soluble myelin oligodendrocyte glycoprotein antigen (MOG) was administered to MOG-TCR transgenic mice (2D2 mice)<sup>16</sup> (Supplementary Fig. 5b). Taken together these data suggest that the generation and the accumulation of T<sub>H</sub>17 cells in the small intestine was not restricted to the CD3-specific antibody treatment, but was a general mechanism following strong T-cell receptor (TCR) stimulation.

We next wanted to identify the molecular signals important for the generation of T<sub>H</sub>17 cells *in vivo* after CD3-specific antibody treatment. Because IL-6 is known to be important for T<sub>H</sub>17 cell generation, we evaluated the importance of this cytokine. *Il6*<sup>−/−</sup> and wild-type mice were treated with CD3-specific antibody. In the *Il6*<sup>−/−</sup> mice, only a very small population of T<sub>H</sub>17 cells (about 2%) could be found by flow cytometry in the small intestine (Supplementary Fig. 6a) and IL-17A was undetectable in the plasma (data not shown). To study the cellular source of IL-6, we treated mice with clodronate-loaded liposomes, which eliminates most macrophages and a significant proportion of dendritic cells compared to PBS-loaded liposomes<sup>13</sup> (Supplementary Fig. 6c). IL-6 plasma levels were greatly reduced in mice treated with clodronate-loaded liposomes compared to control mice after CD3-specific antibody injection (Supplementary Fig. 6d) and a profound reduction in T<sub>H</sub>17 cells was observed (Supplementary Fig. 6b, c). Taken together, these data support the notion that IL-6 secreted by antigen-presenting cells (APCs) is critical for the generation of T<sub>H</sub>17 cells during CD3-specific antibody treatment.

We next analysed the mechanism leading to the specific accumulation of T<sub>H</sub>17 cells in the small intestine, predominantly in the duodenum. T<sub>H</sub>17 cells are known to express the chemokine receptor CCR6 (ref. 17). Whereas CCR6 is relevant in different autoimmune disease models<sup>7,18</sup>, the role of the CCR6/CCL20 axis in immune cell migration to the intestine during tolerance induction has not yet been evaluated. To study that, we analysed the expression of CCR6 on CD4<sup>+</sup> IL-17A–eGFP positive and negative cells (Fig. 2a) and *Ccl20* mRNA expression (Fig. 2b) in the spleen and the gut. CCR6 was mainly expressed in T<sub>H</sub>17 cells from the spleen and the gut 24 h after CD3-specific antibody injection (Fig. 2a). Strikingly, when we performed a time course to measure the mRNA levels of *Ccl20* in different parts of the intestine during CD3-specific antibody treatment, we observed that *Ccl20* was expressed at the highest level in the duodenum in steady state conditions and was selectively further upregulated after CD3-specific antibody treatment (Fig. 2b, c and Supplementary Fig. 7). To test the importance of the CCR6/CCL20 axis for the migration of T<sub>H</sub>17 cells from the periphery to the duodenum, we treated *Ccr6*<sup>−/−</sup> and control mice with CD3-specific antibody. T<sub>H</sub>17 cell number (Fig. 2e) and frequency (Fig. 2d) were strongly reduced in the intestine of the *Ccr6*<sup>−/−</sup> compared to wild-type mice. In general, we did not observe signs of intestinal inflammation in the *Ccr6*<sup>−/−</sup> mice as we did



**Figure 2 | The axis CCR6/CCL20 is essential for the recruitment of T<sub>H</sub>17 cells to the small intestine.** **a**, CCR6 expression 24 h after anti-CD3 treatment. **b**, **c**, *Ccl20* mRNA expression (mean ± s.e.m.; n = 4). S.I., small intestine. **d**, IL-17A expression (gated on CD4<sup>+</sup>TCRβ<sup>+</sup> events) as measured by intracellular cytokine staining. **e**, T<sub>H</sub>17 cell numbers in different organs (mean ± s.d.; n = 5). LN, lymph node; MLN, mesenteric lymph node. **f**, *Ccl20* mRNA expression in duodenum of wild-type (WT), *Il17a*<sup>−/−</sup> and *Il17ra*<sup>−/−</sup> mice (mean ± s.e.m.; n = 4). **g**, *Ccl20* mRNA levels of epithelial and haematopoietic cells isolated from the small intestine. EC, epithelial cells. Panels **b**, **d**–**g** show results 100 h after the first anti-CD3 injection. Data are representative of at least three independent experiments.

in wild-type controls after CD3-specific antibody treatment (data not shown). Interestingly, we detected a higher number of T<sub>H</sub>17 cells in the spleen and lymph nodes of *Ccr6*<sup>−/−</sup> mice when compared to control animals (Fig. 2e). This increase was accompanied by splenomegaly and enlargement of lymph nodes (data not shown), indicating that CCR6 does not have a major role in the generation and expansion of T<sub>H</sub>17 cells. In conclusion, CCR6 seems to be essential for the migration of T<sub>H</sub>17 cells to the small intestine after CD3-specific antibody treatment, and the intestinal inflammation is dependent on this migration. Thus our data indicate that T<sub>H</sub>17 cell migrate to the small intestine leading to intestinal inflammation and damage. However, we cannot exclude that a proliferation of gut resident T<sub>H</sub>17 cells also contributes to the observed phenomenon.

To evaluate the contribution of IL-17A and IL-17F (T<sub>H</sub>17 signature cytokines) in the induction of CCL20 expression in the duodenum, we treated *Il17a*<sup>−/−</sup> or *Il17ra*<sup>−/−</sup> mice with CD3-specific antibody. We found decreased levels of CCL20 in the *Il17a*<sup>−/−</sup> and the *Il17ra*<sup>−/−</sup> mice versus the controls after CD3-specific antibody treatment (Fig. 2f), indicating that IL-17 signalling has a major role in the induction of CCL20 in the duodenum. We next studied the cellular source of CCL20. *Ccl20* mRNA was only detectable in the intestinal epithelial cells in untreated mice. Treatment with CD3-specific antibody led to a

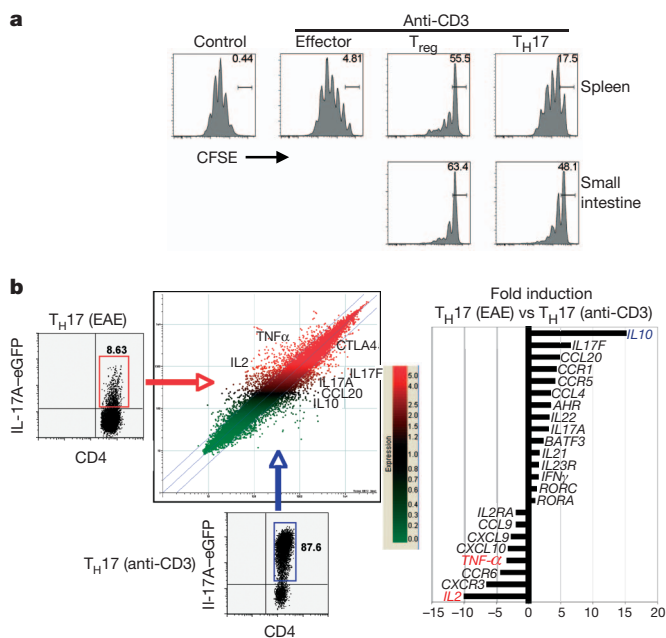
further upregulation of *Ccl20* mRNA by the epithelial cells. Additionally, the CD4<sup>+</sup> T cells present in the small intestine after CD3-specific antibody treatment, most of which were T<sub>H</sub>17 cells, expressed high levels of *Ccl20* mRNA (Fig. 2g). In conclusion, T<sub>H</sub>17 cells via IL-17A and IL-17F production directly upregulate CCL20 production by the intestinal epithelial cells, which then leads to the subsequent recruitment of CCR6<sup>+</sup> T<sub>H</sub>17 cells, which also produce CCL20.

Of note, the intestinal inflammation after CD3-specific antibody treatment was transient and 100% of the mice recovered. To understand better the mechanisms underlying this process, we first assessed apoptosis of T<sub>H</sub>17 cells in the small intestine but we did not detect a significant number of apoptotic cells (data not shown). When we studied the *in vivo* proliferation capacity of CD4<sup>+</sup>TCRβ<sup>+</sup> T cells from the CD3-specific antibody-treated animals, we found that T<sub>H</sub>17 cells from the duodenum were actively proliferating (Supplementary Figs 8 and 9a, b). Using IL17A-eGFP × FoxP3-mRFP double reporter mice (monomeric red fluorescent protein was inserted in the *foxp3* locus) we determined that CD4<sup>+</sup>IL-17A<sup>+</sup> T cells were proliferating at a higher rate than CD4<sup>+</sup>IL-17A<sup>-</sup> T cells in the duodenum (Supplementary Figs 8 and 9). Using two-photon laser-scanning microscopy, we found that the T<sub>H</sub>17 cells in the duodenum did not show the typical behaviour of an apoptotic T cell, conversely, they behaved like activated T cells in terms of their pattern of speed and direction of migration (Supplementary Video). Taken together these data indicate that T<sub>H</sub>17 cells do not die in the small intestine, but are rather actively proliferating.

In line with previous publications<sup>8</sup> we found that CD3-specific antibody treatment caused diarrhoea, oedema, inflammation and tissue destruction in the small intestine (Supplementary Fig. 10a, b), which correlated with the recruitment of T<sub>H</sub>17 cells. However, the intestinal pathology was only transient and mice fully recovered. We therefore began to investigate the fate of T<sub>H</sub>17 cells in the small intestine. Interestingly, we found a fraction of T<sub>H</sub>17 cells in the intestinal lumen

of the CD3-specific antibody-treated mice (Supplementary Fig. 10c, d). Given the severe inflammation, diarrhoea and tissue damage (Supplementary Fig. 10a, b), it is most likely that these cells were passively washed out, although an active mechanism cannot be excluded. Considering that the remaining T<sub>H</sub>17 cells in the duodenum were actively proliferating, but the intestinal pathology was only transient, we were curious about the functional capabilities of these cells. Surprisingly, we found that the remaining T<sub>H</sub>17 cells in the duodenum were able to suppress proliferation of responder T cells *in vitro* (Fig. 3a). To study the molecular properties of these suppressive T<sub>H</sub>17 cells (which we refer from now on as rT<sub>H</sub>17 cells), we performed a genome-wide transcriptional profiling assay (Fig. 3b). We compared the gene expression pattern of rT<sub>H</sub>17 cells from CD3-specific antibody-treated mice and genes expressed by pro-inflammatory T<sub>H</sub>17 cells that were harvested from the central nervous system of EAE-induced mice. The signature genes of T<sub>H</sub>17 cells, like *Rorc*, *Rora*, *Il17a*, *Il22* or *Il23r*, were similarly expressed between both types of T<sub>H</sub>17 cells. Also the activation status of these cells seemed to be similar, because activation markers such as CD69, CD25 and CD44 were equally expressed. However, we found that the rT<sub>H</sub>17 cells from the CD3-specific antibody-treated mice showed a non-inflammatory gene expression profile compared to pro-inflammatory T<sub>H</sub>17 cells isolated from the central nervous system of EAE-induced mice. Notably, the expression levels of *Tnf-α* and *Il-2*, two cytokines with clear pro-inflammatory roles<sup>19,20</sup>, were greatly reduced in the rT<sub>H</sub>17 cells from the small intestine. In contrast, these cells expressed high levels of IL-10, a cytokine with potent anti-inflammatory activities<sup>21</sup> (Supplementary Fig. 11b and Fig. 3b). These data are supported by a previous report showing that *in-vitro*-generated non-pathogenic T<sub>H</sub>17 cells are able to express IL-10 (ref. 22). To evaluate the molecular mechanisms involved in the suppressive function of the rT<sub>H</sub>17 cells, different molecules were blocked in an *in vitro* suppression assay using monoclonal antibodies (Supplementary Fig. 11a). The suppressive capacity of the rT<sub>H</sub>17 cells was partially dependent on IL-10, CTLA-4 and TGF-β. Blocking all three pathways resulted in a lack of suppression by the rT<sub>H</sub>17 cells. T<sub>H</sub>17 cells isolated from the spleen showed an intermediate phenotype. They exhibited a limited capacity to suppress the proliferation of T cells *in vitro* (Fig. 3a), and also produced more TNF-α and IL-2, but less IL-10 compared to T<sub>H</sub>17 cells isolated from the small intestine (Supplementary Fig. 11b). However, because some of the T<sub>H</sub>17 cells in the small intestine downregulated CCR6 (Fig. 2a), it is possible that some rT<sub>H</sub>17 might have migrated back from the small intestine to the spleen. If development of the suppressive capability occurred in the small intestine, then preventing the migration of the T<sub>H</sub>17 cells to that site should prevent the development of these tolerogenic cells. To test this hypothesis, we analysed T<sub>H</sub>17 cells isolated from the spleen of *Ccr6*<sup>-/-</sup> mice, because we showed already that these T<sub>H</sub>17 cells are unable to migrate to the small intestine. Consistent with the hypothesis, *Ccr6*<sup>-/-</sup> T<sub>H</sub>17 cells in the spleen showed high TNF-α production, failed to suppress T-cell proliferation *in vitro*, and were even proinflammatory, causing inflammatory bowel disease *in vivo* upon transfer into a lymphopenic host (Supplementary Fig. 12 a–d). These data indicate that proinflammatory T<sub>H</sub>17 cells do indeed acquire their suppressive phenotype in the small intestine.

To confirm our findings in an animal disease model, EAE-induced mice were treated with CD3-specific antibody. In line with a previous publication<sup>23</sup> we observed a protective effect when the treatment was administered during the course of the disease (Supplementary Fig. 13a). More importantly, we demonstrated that T<sub>H</sub>17 cells were recruited to the duodenum of the CD3-specific antibody treated animals and these mice had strongly reduced numbers of T<sub>H</sub>17 cells in the central nervous system (data not shown). Using a MOG-specific tetramer, we determined that a significant percentage of T<sub>H</sub>17 cells in the duodenum were antigen-specific (Supplementary Fig. 13b), demonstrating that MOG-specific T<sub>H</sub>17 cells were recruited to the duodenum following CD3-specific antibody treatment. In contrast the frequency

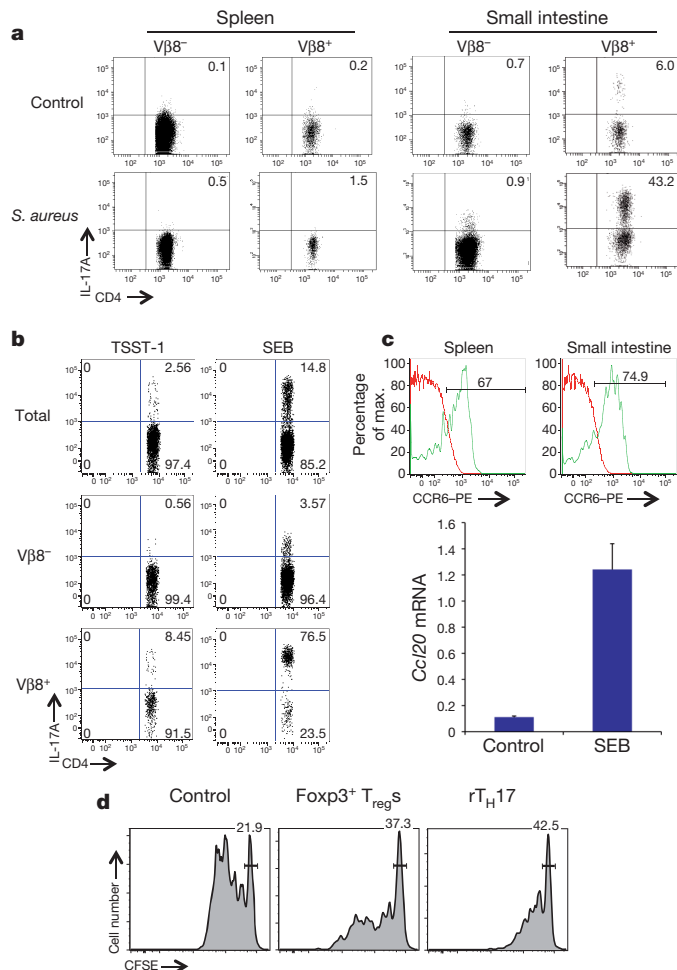


**Figure 3 | Functional and molecular characterization of rT<sub>H</sub>17 cells.** **a**, Suppression assay was performed using eGFP<sup>+</sup>mRFP<sup>+</sup>CD4<sup>+</sup> (Effector), eGFP<sup>+</sup>mRFP<sup>+</sup>CD4<sup>+</sup> (Tregs) or eGFP<sup>+</sup>mRFP<sup>+</sup>CD4<sup>+</sup> (T<sub>H</sub>17) cells sorted from spleen or small intestine. (Bar represents undivided CFSE-labelled CD4<sup>+</sup>CD25<sup>+</sup> responder T cells). Data are representative of six independent experiments. **b**, Gene expression analysis comparing T<sub>H</sub>17 cells (eGFP<sup>+</sup>mRFP<sup>+</sup>CD4<sup>+</sup>) from central nervous system at day 17 after EAE induction versus T<sub>H</sub>17 cells isolated from the small intestine of anti-CD3 treated IL-17A-eGFP × Foxp3-mRFP double reporter mice.



of MOG-tetramer-positive  $T_H17$  cells was much lower in other organs of EAE-induced mice, which had not been treated with CD3-specific antibody (data not shown). This is evidence against a general increase in MOG-specific  $T_H17$ . Therefore our results show that antigen-specific  $T_H17$  cells, with proinflammatory properties, generated in the periphery can be redirected to the small intestine. To confirm that  $rT_H17$  cells isolated from the small intestine of CD3-specific antibody-treated mice are indeed *in vivo* immune-suppressive we tested their suppressive capacity in an EAE transfer model. We co-transferred MOG-specific *in-vitro*-differentiated  $T_H17$  either alone or together with MOG-specific  $rT_H17$  cells isolated from the small intestine of CD3-specific antibody-treated 2D2 transgenic mice. Strikingly, we found that  $rT_H17$  cells were able to completely suppress the development of EAE in these transfer experiments (Supplementary Fig. 13c, d), indicating that the  $rT_H17$  cells are indeed stable in terms of their immune suppressive function.

As mentioned above CD3-specific antibody treatment is already used in clinical trials<sup>9,10</sup>, and we therefore aimed to confirm our results using teplizumab (hOKT3 $\gamma$ 1(Ala-Ala)), one CD3-specific antibody used in these trials. To that end we used a humanized mouse system: we reconstituted Balb/c Rag-2<sup>-/-</sup> $\gamma$ c<sup>-/-</sup> double knockout mice with human peripheral blood mononuclear cells. Two weeks after the transfer we treated these mice with either OKT-3, an FcR-binding



**Figure 4** |  $T_H17$  cells are recruited to the small intestine during sepsis. **a**, **b**, IL-17A-eGFP expression is shown (gated on CD4<sup>+</sup>TCRβ<sup>+</sup> events). Mice were injected with *Staphylococcus aureus* (**a**) or SEB and TSST-1 (**b**). **c**, CCR6 expression 24 h after the first SEB injection (top). *Ccl20* mRNA levels in the small intestine 100 h after the first injection (mean ± s.e.m.; *n* = 4) (bottom). **d**, *In vitro* suppression assay using CD4<sup>+</sup>IL-17A-eGFP<sup>+</sup> cells from the small intestine or CD4<sup>+</sup>Foxp3-mRFP<sup>+</sup> cells from the spleen of SEB-treated mice as suppressor cells. Results are representative of at least two independent experiments.

CD3-specific antibody used in the first human studies, or teplizumab, an FcR non-binding CD3-specific antibody. Strikingly, we found human T cells in the small intestine after treatment with both of these CD3-specific antibodies (Supplementary Fig. 14a, b). The presence of human IL-17A-, IL-10- and CCL20-producing cells in the small intestine in OKT-3- and teplizumab-treated mice was confirmed by real-time PCR (Supplementary Fig. 14c).

Taken together our results obtained in the CD3-specific antibody model suggest that  $T_H17$  cells, by upregulating CCL20 expression in the duodenum via IL-17 signalling, have developed an elegant mechanism to limit the pathogenicity in order to avoid a life-threatening immune response. This predicted in turn that this mechanism should be general to most strong immune responses that result in  $T_H17$  cells.

$T_H17$  cells have a crucial role in controlling different microorganisms *in vivo*<sup>24</sup>. We next investigated whether this mechanism of  $T_H17$  cell control also functions during a strong immune response elicited by pathogenic microorganisms. We first used a murine model of sepsis. We injected *Staphylococcus aureus*, which is one of the most frequent organisms responsible for sepsis in humans<sup>25,26</sup>, intravenously into IL-17A-eGFP reporter mice. Mice were analysed 3 days after the injection, at a time when they displayed severe clinical symptoms of sepsis (weight loss, dehydration, lethargy). Strikingly, we found the highest frequency and number of  $T_H17$  in the small intestine (Fig. 4a). Interestingly, most  $T_H17$  appeared to be TCR Vβ8<sup>+</sup>. The injection of the superantigen SEB (*Staphylococcus aureus* enterotoxin B), which is produced by the bacteria used in these experiments and binds to Vβ8<sup>+</sup> T cells, was sufficient to induce the accumulation of  $T_H17$  in the small intestine just as in the anti-CD3 studies. As a control we injected mice with TSST-1 (toxic shock syndrome toxin 1), a superantigen that does not bind to Vβ8 and is not produced by the bacteria we used. Of note, we observed that the administration of TSST-1 was less effective at inducing the accumulation of  $T_H17$  cells in the small intestine (Fig. 4b). Finally, we could confirm that the  $T_H17$  cells, induced by SEB treatment, expressed CCR6 and that CCL20 is specifically upregulated in the small intestine following SEB treatment (Fig. 4c). Furthermore, while a subpopulation of the  $T_H17$  cells was found in the intestinal lumen (Supplementary Fig. 15), the remaining  $T_H17$  cells demonstrated an immune-suppressive phenotype (Fig. 4d and Supplementary Fig. 16), again comparable to our results obtained in the CD3-specific antibody treatment. Interestingly, it is known that SEB can induce tolerance<sup>27</sup>, which is in line with our results that SEB leads to the generation of  $rT_H17$  cells. Accordingly, we found that SEB and, to a lesser extent, TSST-1 treatment of EAE-induced mice led to the amelioration of disease (data not shown), which is in line with one previous publication<sup>28</sup>.

In addition to anti-bacterial immunity, viruses are the next key class of pathogens to which we must respond, yet contain excessive immunopathology which is commonly the cause of morbidity and mortality<sup>29</sup>. To address such an immune response, we analysed influenza, a viral infection that has devastated human populations. Notably, we again found increased  $T_H17$  cell frequencies in the small intestine in mice infected with influenza A (H1N1) (Supplementary Fig. 17).

In conclusion, we propose a general mechanism that could explain how a pro-inflammatory  $T_H17$  immune response, which is beneficial in clearing infection, but immunopathogenic in excess, can be controlled by the mechanisms we describe here: namely by acquisition of an immune-suppressive phenotype or elimination into the intestinal lumen (Supplementary Fig. 18). These findings and further studies aiming to identify the underlying mechanism of the conversion of pro-inflammatory  $T_H17$  cells into  $rT_H17$  cells may help in designing new strategies to control auto-reactive  $T_H17$  cells in autoimmune diseases like multiple sclerosis.

## METHODS SUMMARY

**Anti-CD3, SEB, TSST-1 treatment and *Staphylococcus aureus* infection.** Mice were injected intraperitoneally three times with either CD3-specific antibody

(clone 2C11, 20 µg per mouse) SEB (50 µg per mouse) or TSST-1 (50 µg per mouse) at 0, 48 and 96 h. Mice were analysed 100 h after the first injection, if not otherwise specified. *Staphylococcus aureus* was injected intravenously ( $1 \times 10^8$  colony-forming units per mouse) in order to induce sepsis. Mice were killed 3 days after the injection.

**Flow cytometric analysis.** Cells were isolated from the organ as indicated. IL-17A-eGFP and CCR6 expression was assessed directly after isolation. When indicated cells were restimulated and intracellular cytokine staining for IL-17A was performed. Numbers in dot-plot quadrants indicate percent cells in each. Cells were gated on CD4<sup>+</sup>TCRβ<sup>+</sup> events.

**Real-time PCR.** *Ccl20* mRNA expression was measured in different tissues as indicated using real time PCR with reverse transcription.

**In vitro suppression assay.** Different suppressor cells were co-cultured with carboxyfluorescein diacetate succinimidyl ester (CFSE)-labelled CD4<sup>+</sup>CD25<sup>-</sup> responder T cells, which were isolated from the spleen of CD45.1 congenic mice. Bar represents undivided CFSE-labelled responder T cells.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 12 May 2010; accepted 19 May 2011.

Published online 17 July 2011.

- Korn, T., Bettelli, E., Oukka, M. & Kuchroo, V. K. IL-17 and Th17 cells. *Annu. Rev. Immunol.* **27**, 485–517 (2009).
- Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201**, 233–240 (2005).
- Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector Th<sub>17</sub> and regulatory T cells. *Nature* **441**, 235–238 (2006).
- Mangan, P. R. *et al.* Transforming growth factor-β induces development of the Th<sub>17</sub> lineage. *Nature* **441**, 231–234 (2006).
- Manel, N., Unutmaz, D. & Littman, D. R. The differentiation of human Th<sub>17</sub> cells requires transforming growth factor-β and induction of the nuclear receptor RORγt. *Nature Immunol.* **9**, 641–649 (2008).
- Yang, L. *et al.* IL-21 and TGF-β are required for differentiation of human Th<sub>17</sub> cells. *Nature* **454**, 350–352 (2008).
- Hirota, K. *et al.* Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and its animal model. *J. Exp. Med.* **204**, 2803–2812 (2007).
- Merger, M. *et al.* Defining the roles of perforin, Fas/FasL, and tumour necrosis factor α in T cell induced mucosal damage in the mouse intestine. *Gut* **51**, 155–163 (2002).
- Chatenoud, L. & Bluestone, J. A. CD3-specific antibodies: a portal to the treatment of autoimmunity. *Nature Rev. Immunol.* **7**, 622–632 (2007).
- Herold, K. C. *et al.* Treatment of patients with new onset type 1 diabetes with a single course of anti-CD3 mAb teplizumab preserves insulin production for up to 5 years. *Clin. Immunol.* **132**, 166–173 (2009).
- Carpenter, P. A. *et al.* Non-Fc receptor-binding humanized anti-CD3 antibodies induce apoptosis of activated human T cells. *J. Immunol.* **165**, 6205–6213 (2000).
- Smith, C. A., Williams, G. T., Kingston, R., Jenkinson, E. J. & Owen, J. J. Antibodies to CD3/T-cell receptor complex induce death by apoptosis in immature T cells in thymic cultures. *Nature* **337**, 181–184 (1989).
- Perruche, S. *et al.* CD3-specific antibody-induced immune tolerance involves transforming growth factor-β from phagocytes digesting apoptotic T cells. *Nature Med.* **14**, 528–535 (2008).
- Chatenoud, L., Primo, J. & Bach, J. F. CD3 antibody-induced dominant self tolerance in overtly diabetic NOD mice. *J. Immunol.* **158**, 2947–2954 (1997).
- Alegre, M. L. *et al.* An anti-murine CD3 monoclonal antibody with a low affinity for Fc gamma receptors suppresses transplantation responses while minimizing acute toxicity and immunogenicity. *J. Immunol.* **155**, 1544–1555 (1995).
- Bettelli, E. *et al.* Myelin oligodendrocyte glycoprotein-specific T cell receptor transgenic mice develop spontaneous autoimmune optic neuritis. *J. Exp. Med.* **197**, 1073–1081 (2003).
- Annuziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J. Exp. Med.* **204**, 1849–1861 (2007).
- Reboldi, A. *et al.* C-C chemokine receptor 6-regulated entry of Th<sub>17</sub> cells into the CNS through the choroid plexus is required for the initiation of EAE. *Nature Immunol.* **10**, 514–523 (2009).
- Brown, S. J. & Mayer, L. The immune response in inflammatory bowel disease. *Am. J. Gastroenterol.* **102**, 2058–2069 (2007).
- Petito, J. M., Streit, W. J., Huang, Z., Butfiloski, E. & Schiffenbauer, J. Interleukin-2 gene deletion produces a robust reduction in susceptibility to experimental autoimmune encephalomyelitis in C57BL/6 mice. *Neurosci. Lett.* **285**, 66–70 (2000).
- Pestka, S. *et al.* Interleukin-10 and related cytokines and receptors. *Annu. Rev. Immunol.* **22**, 929–979 (2004).
- McGeachy, M. J. *et al.* TGF-β and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain Th<sub>17</sub> cell-mediated pathology. *Nature Immunol.* **8**, 1390–1397 (2007).
- Kohm, A. P. *et al.* Treatment with nonmitogenic anti-CD3 monoclonal antibody induces CD4<sup>+</sup> T cell unresponsiveness and functional reversal of established experimental autoimmune encephalomyelitis. *J. Immunol.* **174**, 4525–4534 (2005).
- Khader, S. A., Gaffen, S. L. & Kolls, J. K. Th17 cells at the crossroads of innate and adaptive immunity against infectious diseases at the mucosa. *Mucosal Immunol.* **2**, 403–411 (2009).
- Rodriguez-Creixems, M. *et al.* Bloodstream infections: evolution and trends in the microbiology workload, incidence, and etiology, 1985–2006. *Medicine (Baltimore)* **87**, 234–249 (2008).
- Martin, G. S., Mannino, D. M., Eaton, S. & Moss, M. The epidemiology of sepsis in the United States from 1979 through 2000. *N. Engl. J. Med.* **348**, 1546–1554 (2003).
- Ochi, A., Yuh, K. & Migita, K. Not every superantigen induces tolerance *in vivo*. *Semin. Immunol.* **5**, 57–63 (1993).
- Soos, J. M., Schiffenbauer, J. & Johnson, H. M. Treatment of PL/J mice with the superantigen, staphylococcal enterotoxin B, prevents development of experimental allergic encephalomyelitis. *J. Neuroimmunol.* **43**, 39–43 (1993).
- Chowell, G. *et al.* Severe respiratory disease concurrent with the circulation of H1N1 influenza. *N. Engl. J. Med.* **361**, 674–679 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The authors would like to thank F. Manzo for expert administrative assistance, L. Evangelisti and C. Hughes for generating embryonic stem cells and chimaeric mice, respectively, J. Stein for initial screening of knock-in mice, T. Ferrandino for assistance with the mouse colony, E. Eynon and J. Alderman for managing the mouse program and A. Lin for assistance with Gene Array analysis. We also thank T. Taylor and G. Tokmouline for expert help with the FACS sorting and D. Gonzalez for help with the multiphoton microscopy. We would like to thank J. P. Allison for providing the anti-CTLA-4 antibody, F. Waldron-Lynch and J. S. Pober for providing peripheral blood mononuclear cells, and the NIH Tetramer core facility for providing the tetramers. W.O. was supported by a fellowship from the National Multiple Sclerosis Society. S.H. was supported by the DFG (HU 1714/1-1) and by a James Hudson Brown–Alexander B. Coxe Fellowship. E.E. was supported by the Spanish Ministry of Science postdoctoral fellowship and by a James Hudson Brown–Alexander B. Coxe Fellowship. R.A.F. is an Investigator of the Howard Hughes Medical Institute. The generation of mice for this work was supported by the Transgenic Core of the Yale DERC DK45735 and some of the work supported by a Pilot project from DK45735. This work was also supported by the JDRF.

**Author Contributions** E.E., S.H. and R.A.F. designed the study and wrote the manuscript; N.G. did the *in vitro* suppression assays and the flow analysis for IL-10 expression; A.E.H. and A.M.H. did the two-photon laser-scanning microscopy experiments; T.T. did the immunohistochemistry analysis; W.O. supported the work with key suggestions and by editing the manuscript; E.E. and S.H. did all other *in vitro* and *in vivo* experimental work; Y.Y.W. provided Foxp3–mRFP mice; A.R. did the viral infection experiments; N.V.R. provided clodronate-loaded liposomes, V.K.K. provided 2D2 mice and feedback on the manuscript; Y.I. provided *Il17a*<sup>-/-</sup> mice and feedback on the manuscript; J.K.K. provided the *Il17ra*<sup>-/-</sup> mice and feedback on the manuscript and J.A.B. provided CD3-specific antibodies and feedback on the manuscript; K.C.H. provided teplizumab and key suggestions. E.E. and R.A.F. co-directed the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to E.E. ([enric.esplugues@yale.edu](mailto:enric.esplugues@yale.edu)) or R.A.F. ([richard.flavell@yale.edu](mailto:richard.flavell@yale.edu)).

## METHODS

**Mice.** BALB/c mice (blastocyst donors), CD1 mice (foster mothers), Tet-Cre transgenic mice ("deletor" mice, C57BL/6 background), C57BL/6 mice (B6), C57BL/6.Ly5.1 mice (CD45.1<sup>+</sup>), *IL6*<sup>-/-</sup> mice and *CCR6*<sup>-/-</sup> mice were purchased from The Jackson Laboratories. MOG-transgenic mice (2D2 mice, C57BL/6 background)<sup>30</sup> and Foxp3 reporter mice (FIR mice, C57BL/6 background)<sup>31</sup> were intercrossed with the IL-17A-eGFP reporter mice. We also used *Il17a*<sup>-/-</sup>, *Il17ra*<sup>-/-</sup> and IL-10-eGFP mice (Tiger mice)<sup>32-34</sup>. All mice were kept under specific pathogen-free conditions in the animal care facility at Yale University. The mice were studied at 6–12 week of age. All the experiments were approved by the Institutional Animal Care and Use Committee of Yale University.

**Generation of IL-17A-IRES-eGFP reporter mice.** A BAC clone consisting of *Il17a* genomic DNA derived from C57BL/6 mice was purchased from BacPac (Oakland, CA). An 8-kb BamHI–MluI fragment comprising exons 1, 2 and 3 for the *Il17a* gene was cloned into pEasy-Flox vector adjacent to the thymidine kinase selection marker. The internal ribosome entry site (IRES)–eGFP cassette was linked to a LoxP-flanked neomycin (Neo) selection marker to obtain the IRES-eGFP–Neo cassette. The targeting construct was generated by cloning the IRES-eGFP–Neo cassette into a SacII site between the translation stop codon (UGA) and the polyadenylation signal (A2UA3) of the *Il17a* gene. The targeting construct was linearized by ClaI cleavage and subsequently electroporated into Bruce4 C57BL/6 embryonic stem (ES) cells. Transfected ES cells were selected in the presence of 300 µg ml<sup>-1</sup> G418 and 1 µM ganciclovir. Drug-resistant ES cell clones were screened for homologous recombination by PCR. To obtain chimaeric mice, correctly targeted ES clones were injected into BALB/c blastocysts, which were then implanted into CD1 pseudopregnant foster mothers. Male chimaeras were bred with C57BL/6 to screen for germ-line transmitted offspring. Germ-line transmitted mice were bred with germline Cre transgenic mice (Tet-Cre mice) to remove the neomycin gene. Mice bearing the targeted *Il17a* allele were screened by PCR (*Il17a* knock-in (KI) sense: 5'-CACCAGCGCTGTGTCAAT-3', *Il17a* KI anti-sense: 5'-ACAAACACGAAGCAGTTTGG-3' and *Il17a* IRES: 5'-ACCGGCCTTATTCCAAGC-3').

**Antibodies, tetramers and intracellular cytokine staining.** Anti-CD4 (L3T4), anti-CD62L (MEL-14), anti-CD44 (IM7), anti-CD45.1 (A20) and anti-CD45.2 (104), anti-TCRβ, anti-IL-2, anti-IL-17A, anti-TNFα, anti-Ki-67 and anti-BrdU (5'-bromo-2-deoxyuridine) were purchased from Becton Dickinson Pharmingen. For intracellular cytokine staining, the cells were restimulated with phorbol 12-myristate 13-acetate (PMA) (Sigma, 20 ng ml<sup>-1</sup>) and ionomycin (Sigma, 0.5 µg ml<sup>-1</sup>) for 4 h. GolgiStop (BD Bioscience) was added during the last 3 h of restimulation. After restimulation, the cells were washed and a Ficoll gradient was performed. The cells were fixed with 1% paraformaldehyde (electron microscopy grade) for 10 min on ice. After two washes, cells were incubated with FITC-conjugated anti-GFP antibody (Rockland) and phycoerythrin (PE)-conjugated anti-IL-17A (BD Bioscience) in wash/perm solution (BD Bioscience) for 30 min on ice. Cells were washed twice and resuspended in PBS. Acquisitions were made with a LSRII cytometer (BD Bioscience).

For *ex vivo*-staining with MOG<sub>38-49</sub>/I-A(b)-tetramer-allophycocyanin (APC)-labelled (mouse myelin oligodendrocyte glycoprotein 38-49, "GWYRSPFSRWH", NIH Tetramer Facility), single-cell suspensions were incubated at a density of 10<sup>7</sup> cells ml<sup>-1</sup> with neuraminidase (0.7 µU ml<sup>-1</sup>, neuraminidase type X from *Clostridium perfringens*, Sigma) in serum-free DMEM at 37 °C/10% CO<sub>2</sub> for 25 min before incubation with the I-A(b) multimers (30 µg ml<sup>-1</sup>) in DMEM supplemented with 2% FCS (pH 8.0) at room temperature for 4 h. After washing, cells were stained for 7-AAD (Molecular Probes), CD4 (RM4-5) and TCRβ. hCLIP/I-A(b)-tetramer-APC-labelled was used as a control ("PVSKMRMATPLLMQA", NIH Tetramer Facility). The percentage of tetramer cells was determined in the CD4/TCRβ gate of live (7-AAD<sup>-</sup>) cells. Stained cells were analysed on LSRII cytometer (BD Bioscience) and data were analysed with FlowJo software (Treestar).

**Flow cytometry and FACS sorting.** Collected lymphocytes were treated with ammonium chloride lysis buffer (BioSource International) to remove red blood cells and washed with RPMI containing 10% FBS (Gemini Biological Products). Cells were then stained with a 1:400 dilution of the indicated antibodies together with 10 µg ml<sup>-1</sup> anti-Fc-Receptor blocking antibody (2.4G2, American Type Culture Collection) in PBS containing 2% FBS and then washed twice with PBS. For isolating T cells, CD4<sup>+</sup> T cells were first enriched by magnetic-activated cell sorting beads (Miltenyi Biotec) and then stained with the indicated antibodies. The Becton Dickinson FACSVantage system and MoFlo sorter (DAKO Cytomation) were used for fluorescence detection and cell sorting.

**T<sub>H</sub>17 differentiation *in vitro*.** Splenocytes from IL-17A-IRES-eGFP mice and C57BL/6 mice were incubated with CD4-microbeads and then positively selected through LS columns (Miltenyi Biotec). After enrichment, naive cells (CD4<sup>+</sup> CD25<sup>-</sup> CD62L<sup>hi</sup> CD44<sup>low</sup>) were sorted by FACS as mentioned above. CD4<sup>+</sup> naive T cells were grown for 5 days at 10<sup>6</sup> cells ml<sup>-1</sup> with plate-bound anti-CD3 (5 µg

ml<sup>-1</sup>) and soluble anti-CD28 (2 µg ml<sup>-1</sup>) in medium (Bruff's medium supplemented with 10% FCS, L-glutamine, penicillin and streptomycin) under T<sub>H</sub>17 conditions (TGF-β, IL-6, IL-23, anti-IFN-γ, anti-IL4). IL-17A (eGFP) expression was determined by flow cytometry.

**Multiphoton imaging.** The small intestine (duodenum) from an IL-17A-eGFP × Foxp3-mRFP double reporter mouse treated with CD3-specific antibody was mounted on a glass slide in a chamber consisting of a silicone isolator (20 mm diameter × 0.5 mm, Electron Microscopy Sciences). The tissue was immersed in PBS and covered by a glass coverslip. An Olympus BX50WI microscope equipped with a ×20x, numerical aperture 0.95 Olympus objective and a LaVision TriMScope Multiphoton System controlled by Inspector Software (LaVision Biotec) was used to collect images. For excitation, a Coherent Chameleon Ti:Sapphire laser was tuned to 960 nm. Images of 300 × 300 µm size were recorded at a resolution of 1,024 × 1,024 pixels with 1-µm z-spacing. Emitted light was collected with nondescanned detectors after having passed 435/90, 525/50 and 615/100 nm bandpass filters. Velocity software (Improvision) was used to create three-dimensional image stacks, and QuickTime Pro was used to generate image sequences.

***In vivo* T-cell stimulation and intestinal lymphocyte isolation.** Different mice in C57BL/6 background were injected with anti-CD3 (20 µg, 145 2C11)<sup>34,35</sup> intraperitoneally 1–3 times at an interval of 2 days between injections and killed 4 h after the final injection. For the controls, isotype control or PBS was injected. The intraepithelial lymphocytes (IEL) and lamina propria lymphocytes (LPL) were collected as described with some modifications<sup>34</sup>. In brief, small or large intestines were removed and Peyer's patches were dissected. The first 2 cm of the small intestine were considered as duodenum. Intestines were opened longitudinally and then were cut into strips 1 cm in length. Tissues were washed with Hank's buffered saline and incubated in the presence of 5 mM of EDTA at 37 °C for 30 min. The released cells were loaded onto a Percoll gradient and centrifuged. The cells between 40% and 100% Percoll were collected and used as intestinal epithelial lymphocytes. LPL were collected by digesting gut tissue, which was removed for IEL isolation as described above. The tissue was digested with collagenase IV (100 U, Sigma) at 37 °C for 1 h and loaded onto a Percoll gradient and centrifuged. The cells between 40% and 100% Percoll were collected and used as LPL.

For the lumen content isolation and analysis, mice were anesthetized using isoflurane. One ligation was made after the pylorus and a second one about 4–5 cm distal from the first ligation. A small incision without breaching the vessel proximal to the second ligation was made and 10 ml of pre-warmed PBS (2 ml min<sup>-1</sup>) was injected using a syringe and a 27G1/2 needle. The fluids were collected in a Petri dish placed under the incision proximal to the second ligation. The collected fluids were incubated for 15 min in HBSS/ EDTA and then filtered through a 70 µm cell strainer before FACS analysis.

**Adoptive transfer of CD4<sup>+</sup> T cells.** CD4<sup>+</sup>FoxP3<sup>+</sup> and the CD4<sup>+</sup>FoxP3<sup>-</sup> T cells from the thymus or from the periphery of IL17A-eGFP × Foxp3-mRFP double reporter mice were collected and purified by magnetic-activated cell sorting (MACS; Miltenyi Biotec). After MACS enrichment, total CD4<sup>+</sup>FoxP3<sup>+</sup> and CD4<sup>+</sup>FoxP3<sup>-</sup> T cells were FACS-sorted and 4 × 10<sup>6</sup> T cells were adoptively transferred (intravenously) into sub-lethally irradiated, sex-matched wild-type CD45.1<sup>+</sup> recipient mice. Four weeks after transfer, animals were injected with CD3-specific antibody (20 µg) and the small intestines were recovered and examined for eGFP and mRFP by FACS.

**Immunofluorescence microscopy.** Small intestines were removed from IL-17A-eGFP reporter mice and wild-type littermates after CD3-specific antibody treatment *in vivo*. Small intestines were fixed in 4% paraformaldehyde for 16 h. After two washes with PBS 20% sucrose solution was added. The 20% sucrose solution was replaced 16 h later with 30% sucrose solution. On the next day, the samples were washed and then snap-frozen in OCT and stored at -80 °C. Cryosections were cut at 12 µm on a Leica model CM1850 freezing microtome, applied to Superfrost Plus Gold slides (Fisher Scientific), air-dried, and PAP pen applied (Zymed Laboratories). Sections were blocked for 30 min at ambient temperature with serum-free protein block (Dako Cytomation) and were stained with PE-anti-CD4 (BD) and Alexa488-anti-GFP (Invitrogen) overnight at 4 °C. Samples were washed three times by immersing in PBS for 5 min and then mounted with Prolong gold mounting media with DAPI (Invitrogen). Sections were observed under dark field in independent fluorescence channels using an automated Olympus BX-61 microscope.

**Experimental autoimmune encephalomyelitis.** Mice were immunized subcutaneously with 250 µg of MOG<sub>35-55</sub> (Yale Keck facility) emulsified in CFA (BD Difco). Mice received 400 ng pertussis toxin (PTx, List Biological Laboratories) intraperitoneally at the time of immunization and 48 h later. Mice were checked for clinical symptoms daily, and signs were translated into clinical score as follows: 0, no detectable signs of EAE; 0.5, tail weakness; 1, complete tail paralysis; 2, partial hind limb paralysis; 2.5, unilateral complete hind limb paralysis; 3, complete



bilateral hind limb paralysis; 3.5, complete hind limb paralysis and partial forelimb paralysis; 4, total paralysis of forelimbs and hind limbs (mice with a score above 3.5 to be killed); 5, death. All animal experiments were conducted according to the IACUC policies.

**Cytokine assays.** Cytokines were quantified in plasma by ELISA (TGF- $\beta$ 1, Promega) or by Cytometric Bead Array (IL-6 and IL-17A, BD Bioscience) following the manufacturer's instructions. The plasma was obtained by centrifugation of blood collected on EDTA-coated tubes after cardiac puncture.

**Gene expression analysis.** Total RNA extracted (100 ng; RNeasy, Qiagen) from intestinal rT<sub>H</sub>17 cells (from CD3-specific antibody-treated animals) or from pro-inflammatory T<sub>H</sub>17 cells (from EAE-induced mice) were used to perform a genome-wide transcriptional profiling assay (GeneChip Mouse 1.0 ST Array, Affymetrix). Data was analysed with GeneSpring GX 10 (Agilent Technologies).

**Relative gene expression analysis.** RNA from cells/tissues was isolated with the RNeasy/QIAshredder purification system (Qiagen) in accordance with the manufacturer's protocol. RNA was subjected to reverse transcriptase with Superscript II (Invitrogen) with oligo(dT) primer in accordance with the manufacturer's protocol. cDNA was semi-quantified using commercially available primer/probe sets (Applied Biosystems) and analysed with the  $\Delta\Delta C_t$  (change in cycle threshold) method. All results were normalized to *Hprt* quantified in parallel amplification reactions during each PCR quantification.

**Suppression assays.** CFSE (2  $\mu$ M)-labelled CD4<sup>+</sup>CD25<sup>-</sup> T cells (responder cells) were cultured in 96-well round bottom plates at  $2 \times 10^4$  cells per well with  $10^5$  irradiated APCs (splenocytes MACS-depleted for CD4<sup>+</sup> and CD8<sup>+</sup> T cells) as feeder cells in the presence of  $2 \times 10^4$  cells per well of FACS-sorted CD4<sup>+</sup>IL-17A<sup>+</sup>Foxp3<sup>-</sup> or CD4<sup>+</sup>IL-17A<sup>+</sup>Foxp3<sup>+</sup> T cells. Cell cultures were stimulated with 2  $\mu$ g ml<sup>-1</sup> of anti-CD3 antibody (2C11) in the presence or not of anti-TGF- $\beta$  (1D11), anti-CTLA-4 (9H10) and anti-IL10R. After 4 days, cells were collected, stained and the CFSE signal was analysed by flow cytometry.

**Sepsis induced by infection and superantigen treatment:** *Staphylococcus aureus* (ATCC 14458, SEB<sup>+</sup> TSST-1<sup>-</sup>) was injected intravenously into IL-17A-eGFP reporter mice ( $10^8$  colony-forming units per mouse). Mice were killed 3 days after the injection, at a time when they displayed severe clinical symptoms of sepsis (weight loss, dehydration, lethargy) and the presence of CD4<sup>+</sup>IL17A<sup>+</sup> T cells was tested in different organs (spleen, lymph node, small intestine) using FACS analysis. Similar experiments were done injecting the superantigens SEB and TSST-1. All of them were purchased from Toxin Technology. All superantigens were administered three times (0, 48, 96 h) intraperitoneally at 50  $\mu$ g per mouse.

For the influenza A infection, mice were infected with  $1 \times 10^4$  plaque-forming units of influenza A/PR8 (H1N1) virus via the intranasal route. Infection was

performed by the intranasal application of 50  $\mu$ l virus stock diluted in PBS (or an equal volume of PBS as a control) to mice that had been deeply anesthetized with anafane (Ivesco). Lungs and small intestines were harvested 3 and 5 days after infection for flow cytometry analysis.

**Peripheral blood mononuclear cells isolation and administration.** Human leukocytes were collected by leukapheresis of adult volunteer donors under a protocol approved by the Yale Human Investigations Committee. The peripheral blood mononuclear cells were isolated using Lymphocyte Separation Medium (Cappel) according to the manufacturer's instructions. The cells were stored in 10% DMSO/90% FBS at  $-196^\circ\text{C}$  and were thawed and washed before use. *Rag2*<sup>-/-</sup>  $\times$   *$\gamma$ c*<sup>-/-</sup> double knockout mice were reconstituted with  $5 \times 10^7$  human peripheral blood mononuclear cells by intraperitoneally inoculation 2 weeks before anti-CD3 specific antibody treatment. The number of human T cells (CD45<sup>+</sup>CD4<sup>+</sup>) in the small intestine was evaluated by flow cytometry. Animals demonstrated no signs of graft-vs-host disease. Rare animals that failed to reconstitute with human T cells were, by prior design, excluded from analysis.

**Endoscopic procedure.** Colonoscopy was performed in a blinded fashion for colitis scoring using the Coloview system (Karl Storz, Germany). Briefly: colitis scoring was based on granularity of mucosal surface, stool consistence, vascular pattern, translucency of the colon and fibrin visible (0–3 points for each).

**Statistical analysis.** Where indicated, the Student *t* test for non-paired data and the Mann–Whitney *U* test were used to calculate statistical significance for differences in a particular measurement between different groups. A *P*-value of less than 0.05 was considered significant.

30. Bettelli, E. *et al.* Myelin oligodendrocyte glycoprotein-specific T cell receptor transgenic mice develop spontaneous autoimmune optic neuritis. *J. Exp. Med.* **197**, 1073–1081 (2003).
31. Wan, Y. Y. & Flavell, R. A. Identifying Foxp3-expressing suppressor T cells with a bicistronic reporter. *Proc. Natl Acad. Sci. USA* **102**, 5126–5131 (2005).
32. Nakae, S. *et al.* Antigen-specific T cell sensitization is impaired in IL-17-deficient mice, causing suppression of allergic cellular and humoral responses. *Immunity* **17**, 375–387 (2002).
33. Ye, P. *et al.* Requirement of interleukin 17 receptor signaling for lung CXCR3 chemokine and granulocyte colony-stimulating factor expression, neutrophil recruitment, and host defense. *J. Exp. Med.* **194**, 519–528 (2001).
34. Kamanaka, M. *et al.* Expression of interleukin-10 in intestinal lymphocytes detected by an interleukin-10 reporter knockin *tiger* mouse. *Immunity* **25**, 941–952 (2006).
35. Alegre, M. L. *et al.* An anti-murine CD3 monoclonal antibody with a low affinity for Fc gamma receptors suppresses transplantation responses while minimizing acute toxicity and immunogenicity. *J. Immunol.* **155**, 1544–1555 (1995).

# The crystal structure of a voltage-gated sodium channel

Jian Payandeh<sup>1</sup>, Todd Scheuer<sup>1</sup>, Ning Zheng<sup>1,2</sup> & William A. Catterall<sup>1</sup>

Voltage-gated sodium (Na<sub>V</sub>) channels initiate electrical signalling in excitable cells and are the molecular targets for drugs and disease mutations, but the structural basis for their voltage-dependent activation, ion selectivity and drug block is unknown. Here we report the crystal structure of a voltage-gated Na<sup>+</sup> channel from *Arco bacter butzleri* (NavAb) captured in a closed-pore conformation with four activated voltage sensors at 2.7 Å resolution. The arginine gating charges make multiple hydrophilic interactions within the voltage sensor, including unanticipated hydrogen bonds to the protein backbone. Comparisons to previous open-pore potassium channel structures indicate that the voltage-sensor domains and the S4–S5 linkers dilate the central pore by pivoting together around a hinge at the base of the pore module. The NavAb selectivity filter is short, ~4.6 Å wide, and water filled, with four acidic side chains surrounding the narrowest part of the ion conduction pathway. This unique structure presents a high-field-strength anionic coordination site, which confers Na<sup>+</sup> selectivity through partial dehydration via direct interaction with glutamate side chains. Fenestrations in the sides of the pore module are unexpectedly penetrated by fatty acyl chains that extend into the central cavity, and these portals are large enough for the entry of small, hydrophobic pore-blocking drugs. This structure provides the template for understanding electrical signalling in excitable cells and the actions of drugs used for pain, epilepsy and cardiac arrhythmia at the atomic level.

Electrical signals (termed action potentials) encode and process information within the nervous system and regulate a wide range of physiological processes<sup>1,2</sup>. The voltage-gated ion channels (VGICs) that mediate electrical signalling have distinct functional roles<sup>1,2</sup>. Na<sub>V</sub> channels initiate action potentials. Voltage-gated calcium (Ca<sub>V</sub>) channels initiate processes such as synaptic transmission, muscle contraction and hormone secretion in response to membrane depolarization. Voltage-gated potassium (K<sub>V</sub>) channels terminate action potentials and return the membrane potential to its resting value. Na<sub>V</sub> channels are mutated in inherited epilepsy, migraine, periodic paralysis, cardiac arrhythmia and chronic pain syndromes<sup>3</sup>. These channels are molecular targets of drugs used in local anaesthesia and in the treatment of genetic and sporadic Na<sub>V</sub> channelopathies in the brain, skeletal muscle and heart<sup>4</sup>. The rapid activation, Na<sup>+</sup> selectivity and drug sensitivity of Na<sub>V</sub> channels are unique among VGICs<sup>2</sup>.

VGICs share a conserved architecture in which four subunits or homologous domains create a central ion-conducting pore surrounded by four voltage sensors<sup>5</sup>. The voltage-sensing domain (VSD) is composed of the S1–S4 segments, and the pore module is formed by the S5 and S6 segments with a P-loop between them<sup>5</sup>. The S4 segments place charged amino acids within the membrane electric field that undergo outward displacement in response to depolarization and initiate opening of the central pore<sup>6,7</sup>. Although the architecture of K<sub>V</sub> channels has been established at high resolution<sup>8,9</sup>, the structural basis for rapid, voltage-dependent activation of VGICs remains uncertain<sup>7,9</sup>, and the structures responsible for Na<sup>+</sup>-selective conductance and drug block in Na<sub>V</sub> channels are unknown. The primary pore-forming subunits of Na<sub>V</sub> and Ca<sub>V</sub> proteins in vertebrates are composed of approximately 2,000 amino acid residues in four linked homologous domains<sup>5</sup>. The bacterial NaChBac channel family is an important model for structure–function studies of more complex vertebrate Na<sub>V</sub> and Ca<sub>V</sub> channels<sup>10,11</sup>. NaChBac is a homotetramer, and its pharmacological profile is similar to Na<sub>V</sub> and Ca<sub>V</sub>

channels<sup>10,12</sup>. Bacterial Na<sub>V</sub> channels are highly Na<sup>+</sup> selective, but they can be converted into Ca<sup>2+</sup>-selective forms through simple mutagenesis<sup>13</sup>. The NaChBac family represents the probable ancestor of vertebrate Na<sub>V</sub> and Ca<sub>V</sub> channels. Through analysis of the three-dimensional structure of NavAb from *A. butzleri*, we provide the first insights into the structural basis of voltage-dependent gating, ion selectivity and drug block in Na<sub>V</sub> and Ca<sub>V</sub> channels.

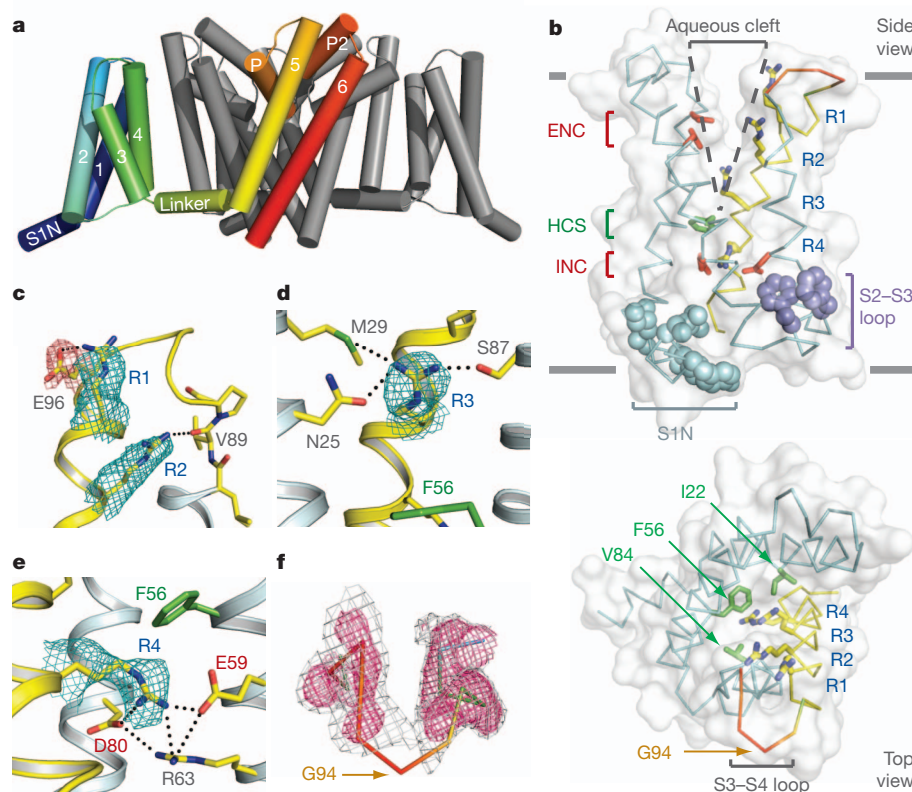
## Structure of NavAb in a membrane environment

NavAb is a member of the NaChBac family and functions as a voltage-gated sodium-selective ion channel (Supplementary Figs 1 and 2). Vertebrate Ca<sub>V</sub> channels require solubilization in digitonin and Na<sub>V</sub> channels require specific lipids to retain function when purified<sup>14,15</sup>. Accordingly, we solubilized NavAb in digitonin, crystallized it in a lipid-based bicelle system, and determined its structure at 2.70 Å resolution (Supplementary Figs 3–6 and Supplementary Table 1). NavAb crystallized as a dimer-of-dimers with 28 lipid molecules bound per tetramer (Supplementary Figs 3 and 6b). Crystal packing indicates a membrane-like environment (Supplementary Fig. 6a). NavAb VSDs interact noncovalently with the pore module of a neighbouring subunit (Fig. 1a), and crystallographic temperature factors highlight their dynamic nature (Supplementary Fig. 6c).

## Structure of the activated voltage sensor

S4 segments in VSDs consist of repeated motifs of a positively charged residue, usually arginine, followed by two hydrophobic residues<sup>5–7</sup>. The R2 and R3 ‘gating charges’ in NavAb are positioned to interact with a conserved extracellular negative-charge cluster (ENC; Fig. 1b), whereas the R4 gating charge interacts with a conserved intracellular negative-charge cluster (INC; Fig. 1b). These structural features, in conjunction with disulphide-locking experiments<sup>16,17</sup>, indicate that the VSDs are in an activated conformation. These ion-pair interactions are expected to stabilize and catalyse S4 movement in the membrane electric field<sup>7,18,19</sup>.

<sup>1</sup>Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.



**Figure 1 | Structure of NavAb and the activated VSD.** **a**, Structural elements in NavAb. One subunit is highlighted (1–6, transmembrane segments S1–S6). The nearest VSD has been removed for clarity. **b**, Side and top views of the VSD illustrating the ENC (red), INC (red), HCS (green), residues of the S1N helix (cyan) and phenylalanines of the S2–S3 loop (purple). S4 segment and gating

charges (R1–R4) are in yellow. **c–e**, Hydrogen bonding of gating charges, dotted lines ( $<3.5$  Å).  $F_o - F_c$  omit maps are contoured over E96 and R1–R4 at 1, 1.5, 2.5 and 1.75 $\sigma$ , respectively. **f**, S3–S4 loop. Coloured according to crystallographic temperature factors of the main chain (blue  $<50$  Å<sup>2</sup> to red  $>150$  Å<sup>2</sup>). An  $F_o - F_c$  omit map is contoured at 1.5 $\sigma$  (grey) and 2.5 $\sigma$  (pink).

Highly conserved Arg 63 in the S2 segment also interacts with R4 and the INC (Fig. 1e), which may stabilize the INC and modulate its electrostatics<sup>20</sup>. NavAb has a spectrum of additional gating charge interactions. R1 interacts with Glu 96, R2 forms a hydrogen bond with the backbone carbonyl of Val 89 in S3, and R3 forms hydrogen bonds with Asn 25 and Met 29 in S1, and Ser 87 in S3 (Fig. 1c–e). This conserved network of hydrogen bonds (Supplementary Fig. 7a) should complement exchange of ion-pair partners and provide a low-energy pathway for S4 movement. The R2–backbone interaction would escape detection in mutagenesis experiments (Fig. 1c) and could have unrecognized significance in the passage of gating charges through the gating pore (Fig. 1b).

The S4 segment in NavAb forms a  $3_{10}$ -helix from R1 to R4. This conformation places all four gating charges in a straight line on one side of S4 (Fig. 1b), such that they could move linearly through the central portion of the gating pore, rather than with a spiral motion<sup>7,17–19</sup>. The S3 segment is a straight  $\alpha$ -helix, and the S3–S4 loop has a dynamic connection to S4 (Fig. 1f). The lack of structural rigidity within the S3–S4 loop (Fig. 1f) indicates that it could move relatively freely in response to large S4 movements during gating.

Our structural analysis reveals further that the S1N helix and S2–S3 loop shield the intracellular surface of the VSD (Fig. 1b and Supplementary Fig. 8). The S2–S3 loop is conserved among VGICs, and two prominent Phe side chains probably stabilize the VSD in the membrane during gating transitions (Fig. 1b and Supplementary Figs 7 and 8)<sup>9</sup>. The S1N-to-S3 region may behave as a modular unit during activation. In contrast to the sheltered intracellular surface of the VSD, a large aqueous cleft extends  $\sim 10$  Å from the extracellular surface into the membrane region above the hydrophobic constriction site (HCS; Fig. 1b). The HCS contains highly conserved residues (Ile 22, Phe 56 and Val 84; Supplementary Fig. 7) that seal the VSD against ion leakage

during S4 movement (Fig. 1b). The NavAb VSD therefore illustrates two important concepts from structure–function studies of  $\text{Na}_v$  channels: a large external vestibule accessible to hydrophilic reagents; and a focused membrane electric field over the intracellular half of the VSD<sup>6,7</sup>.

Despite their separation over one billion years of evolution, the VSDs of NavAb and  $\text{K}_v1.2$  show highly similar conformations (Supplementary Fig. 8a). R4 of NavAb is in an equivalent position to K5 in  $\text{K}_v1.2$  (Supplementary Fig. 8a), the most outward location of K5 during voltage-sensor activation<sup>20</sup>. This observation implies that the NavAb and  $\text{K}_v1.2$  VSDs are both activated.

### The NavAb activation gate is closed

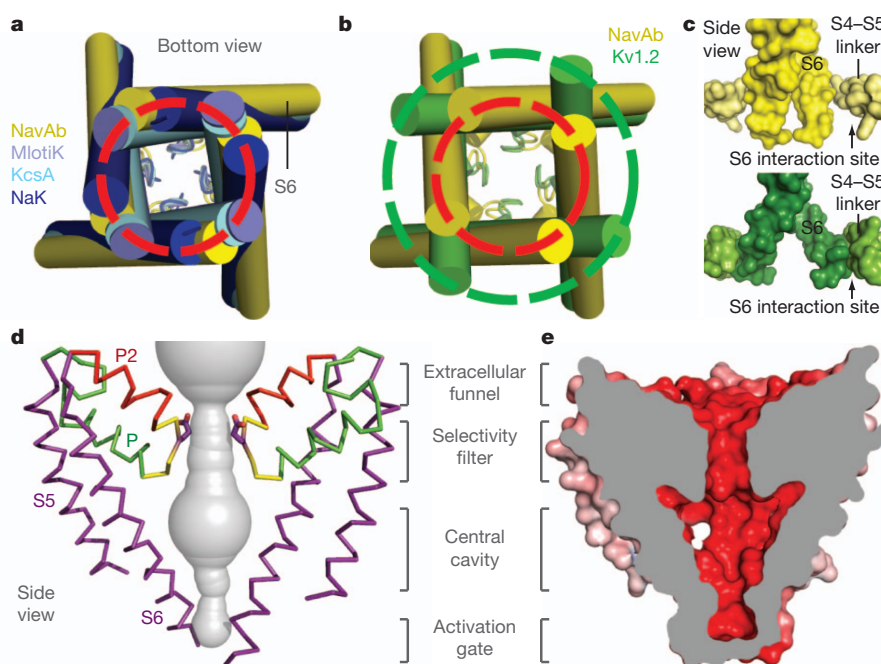
The pore of NavAb is closed, providing the first view of a closed pore in a VGIC (Fig. 2a and Supplementary Fig. 3). Met 221 completely occludes the ion conduction pathway (Supplementary Fig. 4c). The S6 helices of NavAb superimpose well with other closed-pore structures and are distinct from the open-pore  $\text{K}_v1.2$  structure (Fig. 2a, b). A subtle iris-like dilation of the activation gate may be sufficient to open the pore, and the surrounding cuff of S4–S5 linkers may prevent larger pore opening (Fig. 2a–c).

It is surprising to have a closed pore in a VGIC with activated voltage sensors at 0 mV. Our NavAb structures were obtained by introducing a Cys at two locations near the intracellular end of S6 (Ile217Cys or Met221Cys). Evidently, these substitutions allowed us to trap the NavAb channel in the pre-open state previously invoked in kinetic models of VGIC gating (Supplementary Discussion)<sup>21–23</sup>.

### Architecture of the pore and selectivity filter

VGICs are selective for specific cations yet conduct these ions at nearly the rate of free diffusion<sup>2</sup>. Our NavAb structure uncovers a





**Figure 2 | NavAb pore module.** **a**, Pore-lining S6 helices of NavAb (yellow) and the closed MlotiK (PDB code 3BEH), KcsA (PDB code 1K4C) and NaK (PDB code 2AHY) channels. C $\alpha$  locations of Met 221 define a common radius for the closed activation gate (red circle). **b**, Comparison of S6 helices of NavAb and Kv1.2/2.1 (PDB code 2R9R). Dashed circle in red indicates radius of C $\alpha$

atoms of Met 221 in NavAb. **c**, Site for interaction of S6 with S4-S5 linkers (top, NavAb; bottom, Kv1.2/2.1). **d**, Architecture of the NavAb pore. Glu 177 side chains (purple sticks); pore volume is shown in grey. **e**, Electrostatic potential coloured from -10 to 10 kT (red to blue).

basis for the selectivity and high conductance of Na $_v$  channels. The NavAb pore module consists of an outer funnel-like vestibule, a selectivity filter, a central cavity and an intracellular activation gate (Fig. 2d and Supplementary Fig. 4b). The large central cavity in NavAb could easily accommodate a Na $^+$  ion with its first hydration shell and would present a hydrophobic surface over which ions should rapidly diffuse (Fig. 2e and Supplementary Figs 1 and 9). The pore (P)-helices are positioned to stabilize cations in the central cavity through helical-dipole interactions (Fig. 2d and Supplementary Fig. 4b), as suggested for K $^+$  channels<sup>24,25</sup>. Notably, a second pore-helix (P2-helix) forms an extracellular funnel in NavAb (Fig. 2d). This unique P2-helix is not seen in K $^+$  channels and may represent a conserved structural element in the outer vestibule of Na $_v$  and Ca $_v$  channels.

The ion conduction pathway in NavAb is strongly electronegative and the selectivity filter forms the narrowest constriction near the extracellular side of the membrane (Figs 2d, e, 3 and Supplementary Fig. 9). Classic permeation studies suggested a high-field-strength anionic site with dimensions of  $\sim 3.1 \times 5.1$  Å for the selectivity filter in Na $_v$  channels<sup>26,27</sup> and  $5.5 \times 5.5$  Å in Ca $_v$  channels<sup>28</sup>. Mutagenesis studies implicated Glu side chains as key determinants of ion selectivity in these channels<sup>29–33</sup>. In NavAb, the four Glu 177 side chains form a  $\sim 6.5 \times 6.5$  Å scaffold with an orifice of  $\sim 4.6 \times 4.6$  Å defined by van der Waals surfaces (Fig. 3a and Supplementary Fig. 9d). Remarkably, Glu 177 aligns with Glu residues that determine ion selectivity in Na $_v$  and Ca $_v$  channels (Fig. 3e).

The Glu 177 side chains of NavAb are supported by an elaborate architecture (Supplementary Figs 10 and 11). The P-helix ends with the conserved Thr 175, which accepts a hydrogen bond (3.0 Å) from the conserved Trp 179 of a neighbouring subunit (Fig. 3a). This landmark interaction staples together adjacent subunits at the selectivity filter. The residues between Thr 175 and Trp 179 form a tight turn and expose backbone carbonyls of Thr 175 and Leu 176 to conducted ions (Fig. 3b). The Glu 177 side chains form hydrogen bonds with the backbone amides of Ser 180 (2.6 Å) and Met 181 (3.1 Å) from the P2-helix (Fig. 3b and Supplementary Fig. 10). An extensive network of additional interactions (Supplementary Fig. 10), including hydrogen

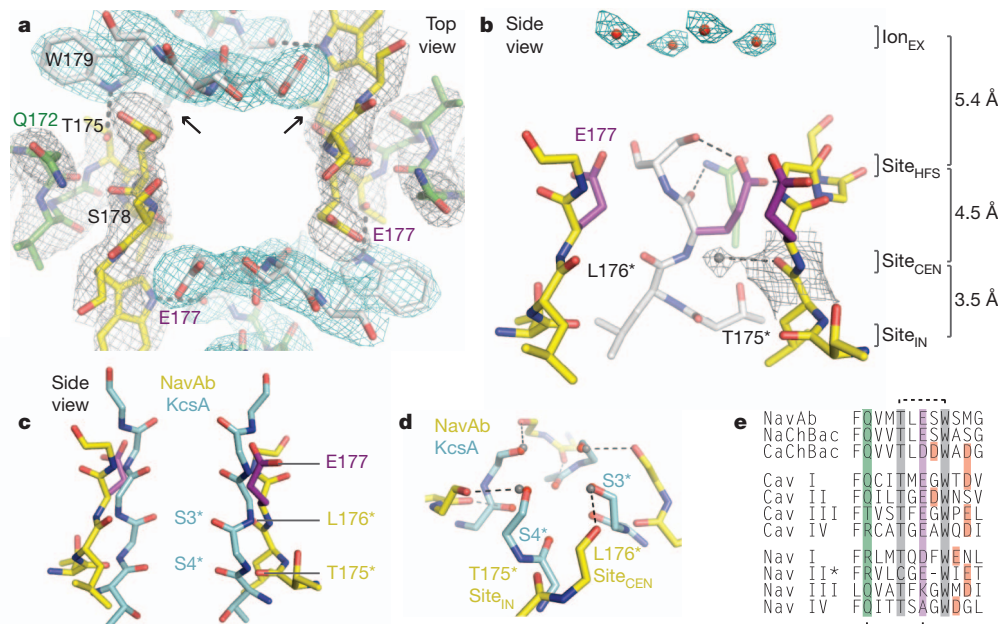
bonds between Gln 172 from the P-helix and the carbonyl of Glu 177 (Fig. 3a, b), further stabilizes the selectivity filter. Owing to the dimer-of-dimers arrangement, the Glu 177 and Ser 178 side chains of NavAb are in two slightly different environments (Fig. 3a and Supplementary Fig. 11), consistent with functional nonequivalence of the corresponding glutamates in Ca $_v$  channels<sup>31–33</sup>.

In agreement with the low affinity of Na $_v$  channels for permeant ions ( $K_d$  for Na $^+$  > 350 mM<sup>34</sup>), no extra density was observed beside the Glu 177 side chains. Instead, strong electron densities were found above Glu 177 at a distance of >4 Å. These densities probably represent cations or solvent molecules (Ion $_{EX}$ ; Fig. 3b) positioned above the selectivity filter by its intense electronegativity (Fig. 2e).

### Ion permeation and selectivity

NavAb represents a prototype for understanding Na $^+$  selectivity and permeation. Analysis of the pore radius indicates that a partially hydrated Na $^+$  ion can be accommodated at the high-field-strength site formed by the Glu 177 side chains (Site $_{HFS}$ ; Fig. 3a, b and Supplementary Fig. 9d). The much narrower K $^+$ -channel filter can fit inside the NavAb selectivity filter (Fig. 3c). Careful inspection of the electron density indicates four well-bound water molecules 2.5 Å from the Leu 176 carbonyls (Site $_{CEN}$ ; Fig. 3b). Remarkably, these four water molecules occupy the same positions as the site 3 carbonyls from K $^+$  channels (Fig. 3c, d)<sup>35</sup>. A distance of 2.5 Å is also found between the backbone carbonyls of Thr 175 from NavAb and the site 4 carbonyls of K $^+$  channels (Fig. 3c, d)<sup>35</sup>. Analogous to other Na $^+$  complexes (Supplementary Fig. 12)<sup>36–40</sup>, a Na $^+$  ion surrounded by a square array of four water molecules could interact with the backbone carbonyls of Leu 176 (Site $_{CEN}$ ) or Thr 175 (Site $_{IN}$ ) (Fig. 3d and Supplementary Fig. 12). Therefore, unlike K $^+$  channels, the NavAb selectivity filter seems to select and conduct Na $^+$  ions in a mostly hydrated form.

The NavAb structure fits closely with Hille's single-ion pore model for Na $_v$  channels, in which a high-field-strength anion partially dehydrates the permeating ion<sup>2,34</sup>. According to Eisenman's theory<sup>41</sup>, a Na $^+$  ion would approach the Site $_{HFS}$  more closely than the larger



**Figure 3 | Structure of the NavAb selectivity filter.** **a**, Top view of the selectivity filter. Symmetry-related molecules are coloured white and yellow; P-helix residues are coloured green. Hydrogen bonds between Thr 175 and Trp 179 are indicated by grey dashes. Electron densities from  $F_o - F_c$  omit maps are contoured at  $4.0\sigma$  (blue and grey) and subtle differences can be appreciated (small arrows). **b**, Side view of the selectivity filter. Glu 177 (purple) interactions with Gln 172, Ser 178 and the backbone of Ser 180 are shown in the far subunit.  $F_o - F_c$  omit map,  $4.75\sigma$  (blue); putative cations or water molecules (red spheres,  $\text{Ion}_{\text{EX}}$ ). Electron density around Leu 176 (grey;  $F_o - F_c$  omit map

$\text{K}^+$  ion, allowing more efficient removal of water and faster permeation (Fig. 3a, b)<sup>34</sup>. A  $\text{Na}^+$  ion could fit in-plane between the Glu 177 side chains, with one side chain coordinating the  $\text{Na}^+$  ion directly and neighbouring Glu 177 side chains acting as hydrogen bond acceptors for two in-plane water molecules<sup>26,27,34</sup>. With two additional waters remaining axial to the ion, this arrangement would approximate trigonal bipyramidal coordination<sup>38</sup>. Because only one Glu 177 side chain engages the permeating ion directly, this transient complex would be inherently asymmetric. When the permeating ion escapes Site<sub>HFS</sub>, full rehydration would occur along the water-lined sites formed by the backbone carbonyls of Leu 176 (Site<sub>CEN</sub>) and Thr 175 (Site<sub>IN</sub>; Fig. 3b, d and Supplementary Fig. 12). Free diffusion then allows the hydrated  $\text{Na}^+$  ion to enter the central cavity and move through the open activation gate into the cytoplasm<sup>34</sup>. The selectivity-filter structure of NavAb concentrates barriers to ion flow into  $\sim 5 \text{ \AA}$  (Fig. 3b and Supplementary Fig. 9d), which should promote high flux rates<sup>34</sup>. This permeation mechanism probably reflects the high free energy of  $\text{Na}^+$  hydration, where further removal of solvating waters would present too high an energy barrier. In sharp contrast,  $\text{K}^+$ -selective channels conduct nearly fully dehydrated  $\text{K}^+$  ions through direct interactions with backbone carbonyls in a long, narrow, multi-ion pore<sup>35,36</sup>. The architectures of the selectivity filters of vertebrate  $\text{Na}_v$  and  $\text{Ca}_v$  channels probably resemble NavAb, and amino acid substitutions within this structural framework must impart  $\text{Na}^+$  versus  $\text{Ca}^{2+}$  selectivity (Supplementary Discussion)<sup>13,29–33</sup>.

### Interaction sites of pore blockers

NavAb provides a foundation to interpret pharmacological mechanisms. From the extracellular side, the Glu 177 side chains of NavAb represent the blocking site in  $\text{Na}_v$  channels for protons and guanidinium moieties of tetrodotoxin and saxitoxin<sup>2,42</sup>, as well as the site where divalent cations and protons bind and block  $\text{Ca}_v$  channels (Fig. 3)<sup>31–33</sup>. From the intracellular side, local anaesthetics, antiarrhythmics and

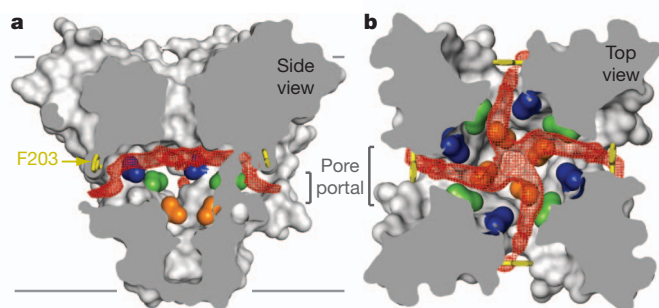
antiepileptic drugs block  $\text{Na}_v$  and  $\text{Ca}_v$  channels<sup>2,4</sup> by entering through the open intracellular mouth of the pore and binding to an overlapping receptor site on the S6 segments<sup>43–45</sup>. Alignment of NavAb S6 segments with vertebrate  $\text{Na}_v$  and  $\text{Ca}_v$  channels reveals a high degree of sequence similarity (Supplementary Fig. 7b), and drug molecules could easily fit into the large central cavity (Fig. 2e and Supplementary Fig. 9). Use-dependent block is enhanced by repetitive opening of the pore to provide drug access<sup>2,46</sup>, and the local anaesthetic etidocaine is an open-channel blocker of NaChBac<sup>12</sup>. The tight seal observed at the intracellular activation gate in NavAb illustrates why pore opening is required for access of large or hydrophilic drugs to the S6 receptor site (Fig. 2 and Supplementary Fig. 4c).

### Fenestrations provide hydrophobic access to pore

Membrane lipids modulate the structure and function of VGICs<sup>8,9,47,48</sup>. However, NavAb presents a completely unexpected type of lipid interaction that has profound implications. The NavAb central cavity reveals four lateral openings leading from the membrane to the lumen of the closed pore (Fig. 4). These fenestrations measure  $\sim 8 \times 10 \text{ \AA}$ , and could become larger depending upon nearby side-chain conformations (Phe 203; Fig. 4). Lipids penetrate through these side portals and lie deep within the central cavity, occluding the ion conduction pathway in NavAb (Fig. 4, red). Because acyl-chain-containing detergents were never used in the preparation of NavAb crystals, these electron densities are assigned as acyl chains of membrane phospholipids. Similar fenestrations were not observed in the open-pore structure of  $\text{K}_v1.2$  (refs 8, 9), raising the possibility that these lipid chains withdraw and the fenestrations close in the open state.

The lateral pore fenestrations in NavAb lead directly to the drug-binding sites within the central cavity and abut residues that are important for drug binding in  $\text{Na}_v$  and  $\text{Ca}_v$  channels (Fig. 4 and Supplementary Fig. 7b)<sup>43,44</sup>. These NavAb portals appear compatible





**Figure 4 | Membrane access to the central cavity in NavAb.** **a**, Side view through the pore module illustrating fenestrations (portals) and hydrophobic access to central cavity. Phe 203 side chains are shown as yellow sticks. Surface representations of NavAb residues aligning with those implicated in drug binding and block: Thr 206, blue; Met 209, green; Val 213, orange. Membrane boundaries, grey lines. Electron density from an  $F_o - F_c$  omit map is contoured at  $2.0\sigma$ . **b**, Top view sectioned below the selectivity filter, coloured as in **a**.

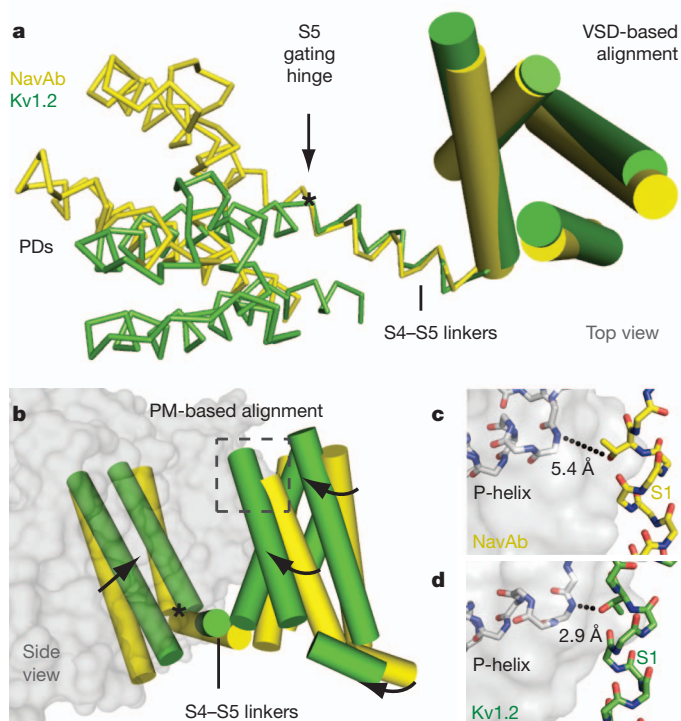
with the passage of small neutral or hydrophobic drugs such as phenytoin<sup>49</sup> and benzocaine<sup>46</sup>, which can gain access to their receptor site in closed channels<sup>2,46</sup>. We propose that pore fenestrations may be directly involved in voltage-dependent drug block according to the ‘modulated receptor model’<sup>46</sup>. Our findings highlight the potential for lipids and other hydrophobic molecules to influence the function of ion channels from the lipid phase of the membrane.

### Structural basis for central pore gating

The domain-swapped arrangement of the VSD around the pore allows the S4–S5 linker to couple S4 movements to activation of VGICs (Fig. 1a)<sup>9</sup>. Kinetic models indicate that all four voltage sensors activate and then the central pore opens in a concerted transition<sup>21–23</sup>. An essential element of this gating model is a state in which all four VSDs have activated but the pore remains closed<sup>21–23</sup>. It is likely that we have captured this pre-open state in our crystals (Supplementary Discussion). NavAb therefore provides a unique opportunity to consider the structural basis for coupling of VSD activation to pore opening.

When activated VSDs of NavAb and Kv1.2 are overlaid (Supplementary Fig. 8a), the S4–S5 linkers superimpose precisely, but the pore domains diverge at the foot of S5 (Fig. 5a). Superposition of the pore domains demonstrates an equivalent displacement of the VSDs (Supplementary Fig. 13). These comparisons lead to a working model for pore opening. First, during activation, the S4–S5 linker and VSD move together as a modular unit (Fig. 5a). Second, a single molecular hinge at the base of S5 mediates the closed-to-open pore transition (Fig. 5a, b). Third, tight structural coupling is maintained between the S5 and S6 segments (Supplementary Fig. 13a). This model suggests that rotation of the VSD and S4–S5 linker as a structural unit pulls the S5–S6 helices outward to open the pore (Fig. 5b and Supplementary Fig. 13b). Because of their tight structural coupling, displacement of the S5–S6 segments from one subunit forces the neighbouring subunits to move similarly, leading to concerted pore opening. During this transition, the amphipathic S4–S5 linker pivots along the plane of the membrane interface (Fig. 5b and Supplementary Figs 7 and 13b). In contrast to Kv1.2, the S6 helices in NavAb have not fully engaged their interaction site on the S4–S5 linker (Fig. 2c), in agreement with the pre-open state of NavAb. A rolling motion of the VSDs around the pore produces displacements up to  $\sim 10$  Å at the intracellular side (Fig. 5b and Supplementary Fig. 13b), which may influence movements of the S1N helix and the conserved S2–S3 loop.

In NavAb, a  $3_{10}$ -helix extends from R1 to R4 (Fig. 1b). In Kv1.2, a  $3_{10}$ -helix encompasses R3 to K5 (equivalent to NavAb R2 to R4), but



**Figure 5 | Model for activation gate opening.** **a**, Superposition of NavAb and Kv1.2/2.1 on the basis of their VSDs (cylinders). PDs, pore domains. **b**, Superposition of NavAb and Kv1.2/2.1 tetrameric pore modules (PM) viewed from the membrane. S5 gating hinge is indicated with an asterisk. Dashed square is enlarged in panels **c** and **d**. **c**, **d**, S1 interaction with P-helix. The distance from the S1 Thr to the P-helix of the neighbouring subunit is 2.9 Å in Kv1.2/2.1, but  $>4.5$  Å in NavAb.

the remaining S4 segment is  $\alpha$ -helical<sup>9</sup>. Conceivably, energy derived from voltage-driven translocation of S4 may be stored in the higher-energy  $3_{10}$ -helix, and then released to help drive pore opening. The VSDs in Kv1.2 are displaced outward ( $\sim 2$  Å) compared to the pre-open NavAb structure (Fig. 5b), which could account for the small gating current associated with concerted pore opening<sup>6</sup>. At the extracellular side of the VSD, an S1 threonine residue hydrogen bonds (2.9 Å) with the P-helix of a neighbouring subunit in Kv1.2 (Fig. 5d), providing a conserved contact point that allows the VSD to perform mechanical work on the pore<sup>50</sup>. The equivalent S1 threonine in NavAb has not yet engaged the P-helix (Fig. 5c). This interaction may therefore represent an essential step in activation gating that has not yet occurred in the pre-open state of NavAb.

### Conclusion

The structure of NavAb provides key insights into the molecular basis of voltage sensing, ion conductance and voltage-dependent gating in a historic class of ion channels<sup>1,2</sup>. A new network of interactions within the VSD appears well positioned to catalyse gating charge movements during activation. Our model for electromechanical coupling reveals a rolling motion of the VSD and its connecting S4–S5 linker around the pore. The NavAb selectivity filter illustrates the basis for selective  $\text{Na}^+$  conductance through a water-lined pore featuring a high-field-strength anionic site. Lastly, hydrophobic access from the membrane phase has been uncovered as a potentially important pathway for drug binding and modulation of VGICs.

### METHODS SUMMARY

NavAb was expressed in insect cells and purified using anti-Flag resin and size-exclusion chromatography, reconstituted into DMPC:CHAPSO bicelles, and crystallized over an ammonium sulphate solution containing 0.1 M Na-citrate,



pH 4.75. Cysteine mutants were complexed with mercury to obtain initial experimental phases. A single anomalous dispersion (SAD) data set from a mercury-free SeMet-substituted protein crystal expedited model building. Standard crystallographic refinement procedures and structural analyses were performed. Electrophysiological experiments on NavAb were performed in tsA-201 cells using standard protocols.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 January; accepted 27 May 2011.**

**Published online 10 July; corrected 21 July 2011 (see full-text HTML version for details).**

- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
- Hille, B. *Ion Channels of Excitable Membranes* 3rd edn (Sinauer Associates, 2001).
- Ryan, D. P. & Ptacek, L. J. Episodic neurological channelopathies. *Neuron* **68**, 282–292 (2010).
- Catterall, W. A. Common modes of drug action on Na<sup>+</sup> channels: local anesthetics, antiarrhythmics and anticonvulsants. *Trends Pharmacol. Sci.* **8**, 57–65 (1987).
- Yu, F. H. & Catterall, W. A. The VGL-channome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Sci. STKE* **2004**, re15 (2004).
- Bezanilla, F. The action potential: from voltage-gated conductances to molecular structures. *Biol. Res.* **39**, 425–435 (2006).
- Catterall, W. A. Ion channel voltage sensors: structure, function, and pathophysiology. *Neuron* **67**, 915–928 (2010).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K<sup>+</sup> channel. *Science* **309**, 897–903 (2005).
- Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K<sup>+</sup> channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
- Ren, D. *et al.* A prokaryotic voltage-gated sodium channel. *Science* **294**, 2372–2375 (2001).
- Koishi, R. *et al.* A superfamily of voltage-gated sodium channels in bacteria. *J. Biol. Chem.* **279**, 9532–9538 (2004).
- Zhao, Y., Scheuer, T. & Catterall, W. A. Reversed voltage-dependent gating of a bacterial sodium channel with proline substitutions in the S6 transmembrane segment. *Proc. Natl Acad. Sci. USA* **101**, 17873–17878 (2004).
- Yue, L., Navarro, B., Ren, D., Ramos, A. & Clapham, D. E. The cation selectivity filter of the bacterial sodium channel, NaChBac. *J. Gen. Physiol.* **120**, 845–853 (2002).
- Curtis, B. M. & Catterall, W. A. Reconstitution of the voltage-sensitive calcium channel purified from skeletal muscle transverse tubules. *Biochemistry* **25**, 3077–3083 (1986).
- Feller, D. J., Talvenheimo, J. A. & Catterall, W. A. The sodium channel from rat brain. Reconstitution of voltage-dependent scorpion toxin binding in vesicles of defined lipid composition. *J. Biol. Chem.* **260**, 11542–11547 (1985).
- DeCaen, P. G., Yarov-Yarovoy, V., Zhao, Y., Scheuer, T. & Catterall, W. A. Disulfide locking a sodium channel voltage sensor reveals ion pair formation during activation. *Proc. Natl Acad. Sci. USA* **105**, 15142–15147 (2008).
- DeCaen, P. G., Yarov-Yarovoy, V., Sharp, E. M., Scheuer, T. & Catterall, W. A. Sequential formation of ion pairs during activation of a sodium channel voltage sensor. *Proc. Natl Acad. Sci. USA* **106**, 22498–22503 (2009).
- Catterall, W. A. Molecular properties of voltage-sensitive sodium channels. *Annu. Rev. Biochem.* **55**, 953–985 (1986).
- Yarov-Yarovoy, V., Baker, D. & Catterall, W. A. Voltage sensor conformations in the open and closed states in ROSETTA structural models of K<sup>+</sup> channels. *Proc. Natl Acad. Sci. USA* **103**, 7292–7297 (2006).
- Tao, X., Lee, A., Limapichat, W., Dougherty, D. A. & MacKinnon, R. A gating charge transfer center in voltage sensors. *Science* **328**, 67–73 (2010).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Shaker potassium channel gating. III: Evaluation of kinetic models for activation. *J. Gen. Physiol.* **103**, 321–362 (1994).
- Kuzmenkin, A., Bezanilla, F. & Correa, A. M. Gating of the bacterial sodium channel, NaChBac: voltage-dependent charge movement and gating currents. *J. Gen. Physiol.* **124**, 349–356 (2004).
- Zhao, Y., Yarov-Yarovoy, V., Scheuer, T. & Catterall, W. A. A gating hinge in Na<sup>+</sup> channels: a molecular switch for electrical signaling. *Neuron* **41**, 859–865 (2004).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **280**, 69–77 (1998).
- Jogini, V. & Roux, B. Electrostatics of the intracellular vestibule of K<sup>+</sup> channels. *J. Mol. Biol.* **354**, 272–288 (2005).
- Hille, B. The permeability of the sodium channel to organic cations in myelinated nerve. *J. Gen. Physiol.* **58**, 599–619 (1971).
- Hille, B. The permeability of the sodium channel to metal cations in myelinated nerve. *J. Gen. Physiol.* **59**, 637–658 (1972).
- McCleskey, E. W. & Almers, W. The Ca channel in skeletal muscle is a large pore. *Proc. Natl Acad. Sci. USA* **82**, 7149–7153 (1985).
- Heinemann, S. H., Terlau, H., Stuhmer, W., Imoto, K. & Numa, S. Calcium channel characteristics conferred on the sodium channel by single mutations. *Nature* **356**, 441–443 (1992).
- Favre, I., Moczydlowski, E. & Schild, L. On the structural basis for ionic selectivity among Na<sup>+</sup>, K<sup>+</sup>, and Ca<sup>2+</sup> in the voltage-gated sodium channel. *Biophys. J.* **71**, 3110–3125 (1996).
- Yang, J., Ellinor, P. T., Sather, W. A., Zhang, J. F. & Tsien, R. W. Molecular determinants of Ca<sup>2+</sup> selectivity and ion permeation in L-type Ca<sup>2+</sup> channels. *Nature* **366**, 158–161 (1993).
- Ellinor, P. T., Yang, J., Sather, W. A., Zhang, J. F. & Tsien, R. W. Ca<sup>2+</sup> channel selectivity at a single locus for high-affinity Ca<sup>2+</sup> interactions. *Neuron* **15**, 1121–1132 (1995).
- Chen, X. H., Bezprozvanny, I. & Tsien, R. W. Molecular basis of proton block of L-type Ca<sup>2+</sup> channels. *J. Gen. Physiol.* **108**, 363–374 (1996).
- Hille, B. Ionic selectivity, saturation, and block in sodium channels. A four-barrier model. *J. Gen. Physiol.* **66**, 535–560 (1975).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. Energetic optimization of ion conduction rate by the K<sup>+</sup> selectivity filter. *Nature* **414**, 37–42 (2001).
- Ye, S., Li, Y. & Jiang, Y. Novel insights into K<sup>+</sup> selectivity from high-resolution structures of an open K<sup>+</sup> channel pore. *Nature Struct. Mol. Biol.* **17**, 1019–1023 (2010).
- Alam, A. & Jiang, Y. Structural analysis of ion selectivity in the NaK channel. *Nature Struct. Mol. Biol.* **16**, 35–41 (2009).
- Doi, M. *et al.* Caged and clustered structures of endothelin inhibitor BQ123, cyclo[D-Trp-D-Asp<sup>1</sup>-Pro-D-Val-Leu<sup>2</sup>]-Na<sup>+</sup>, forming five and six coordination bonds between sodium ions and peptides. *Acta Crystallogr. D* **57**, 628–634 (2001).
- Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
- Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.* **273**, 171–182 (1997).
- Eisenman, G. & Horn, R. Ionic selectivity revisited: the role of kinetic and equilibrium processes in ion permeation through channels. *J. Membr. Biol.* **76**, 197–225 (1983).
- Noda, M., Suzuki, H., Numa, S. & Stuhmer, W. A single point mutation confers tetrodotoxin and saxitoxin insensitivity on the sodium channel II. *FEBS Lett.* **259**, 213–216 (1989).
- Hockerman, G. H., Peterson, B. Z., Johnson, B. D. & Catterall, W. A. Molecular determinants of drug binding and action on L-type calcium channels. *Annu. Rev. Pharmacol. Toxicol.* **37**, 361–396 (1997).
- Ragsdale, D. S., McPhee, J. C., Scheuer, T. & Catterall, W. A. Molecular determinants of state-dependent block of Na<sup>+</sup> channels by local anesthetics. *Science* **265**, 1724–1728 (1994).
- Ragsdale, D. S., McPhee, J. C., Scheuer, T. & Catterall, W. A. Common molecular determinants of local anesthetic, antiarrhythmic, and anticonvulsant block of voltage-gated Na<sup>+</sup> channels. *Proc. Natl Acad. Sci. USA* **93**, 9270–9275 (1996).
- Hille, B. Local anesthetics: hydrophilic and hydrophobic pathways for the drug-receptor reaction. *J. Gen. Physiol.* **69**, 497–515 (1977).
- Oliver, D. *et al.* Functional conversion between A-type and delayed rectifier K<sup>+</sup> channels by membrane lipids. *Science* **304**, 265–270 (2004).
- Delmas, P., Coste, B., Gamper, N. & Shapiro, M. S. Phosphoinositide lipid second messengers: new paradigms for calcium channel modulation. *Neuron* **47**, 179–182 (2005).
- Morello, R. S., Begenisich, T. & Yeh, J. Z. Determination of the active form of phenytoin. *J. Pharmacol. Exp. Ther.* **230**, 156–161 (1984).
- Lee, S. Y., Banerjee, A. & MacKinnon, R. Two separate interfaces between the voltage sensor and pore are required for the function of voltage-dependent K<sup>+</sup> channels. *PLoS Biol.* **7**, e47 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Hille for comments on a draft of the manuscript and members of the N.Z. and W.A.C. groups for their support throughout this project. We are grateful to investigators who provided genomic DNA and the beamline staff at the Advanced Light Source (BL8.2.1 and BL8.2.2) for their assistance during data collection. J.P. acknowledges support from a Canadian Institutes of Health Research fellowship and the encouragement of E. Payandeh. This work was supported by grants from the National Institutes of Health (R01 NS15751 and U01 NS058039 to W.A.C.) and by the Howard Hughes Medical Institute (N.Z.).

**Author Contributions** N.Z. and W.A.C. are co-senior authors. J.P., N.Z. and W.A.C. conceived and J.P. conducted the protein purification and crystallization experiments. J.P. and N.Z. determined and analysed the structures of NavAb. J.P. and T.S. performed functional studies of NavAb. J.P., N.Z. and W.A.C. wrote the manuscript.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 3RVY, 3RVZ and 3RW0. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.Z. (nzheng@uw.edu) and W.A.C. (wcatt@uw.edu).

## METHODS

**Protein expression and purification.** After exploring traditional expression approaches in *Escherichia coli*<sup>51</sup>, the NavAb channel from *A. butzleri* was cloned into the pFASTBac-Dual vector behind the polyhedron promoter using the BamHI and NotI restriction sites preceded by an N-terminal Flag tag. Recombinant baculovirus were generated using the Bac-to-Bac system (Invitrogen) and insect cells were infected for large-scale protein production. Cells were harvested 72 h post-infection and resuspended in 50 mM Tris pH 8.0, 200 mM NaCl (Buffer A) supplemented with protease inhibitors and DNase. After sonication, digitonin (EMD Biosciences) was added to 1% and solubilization was carried out for 1–2 h at 4 °C. After centrifugation, clarified supernatant was gently agitated with anti-Flag M2-agarose resin (Sigma) pre-equilibrated with Buffer B (Buffer A supplemented with 0.12% digitonin) for 1–2 h at 4 °C. Flag resin was collected in a column by gravity flow, washed with ten column volumes of Buffer B, and eluted with two column volumes of Buffer B supplemented with 0.1 mg ml<sup>-1</sup> Flag peptide. The eluate was passed over a Superdex 200 column (GE Healthcare) in 10 mM Tris pH 8, 100 mM NaCl and 0.12% digitonin and peak fractions containing NavAb were concentrated using a Vivaspinn (30K MWKO) centrifugal device. Site-directed mutagenesis was performed using the standard QuikChange protocol (Stratagene) and all constructs were confirmed by DNA sequencing. Selenomethionine-labelled proteins were expressed using established protocols<sup>52</sup>, except cells were washed and starved for methionine at 8 h after infection, followed by SeMet (Anatrace) supplementation at 12 h after infection. SeMet-labelled proteins were purified as described earlier. **Heavy atom screening and labelling.** During our efforts to identify useful derivatives for crystallographic phasing, we ultimately turned to the fluorescence detection of heavy atom labelling (FD-HAL) method<sup>53</sup>. Over thirty NavAb single-site cysteine mutations were rapidly screened using the FD-HAL method, and many of these mutant proteins were subsequently crystallized, presumably as covalent mercury-channel complexes. The NavAb(Ile217Cys) and NavAb(Met221Cys) mutants that yielded useful single anomalous dispersion (SAD) data sets were prepared as follows: proteins were purified as described earlier and concentrated to ~1 mg ml<sup>-1</sup>; HgCl<sub>2</sub> was added to a final concentration of 10 mM and the mixture was incubated at room temperature (22 °C) for 2 h. The protein buffer was subsequently exchanged (into mercury-free buffer) through five rounds of concentration and dilution using Vivaspinn (30K MWKO) centrifugal devices. Following structure determination, it became apparent that Met 221 lines the narrowest portion of the closed NavAb pore.

**NavAb crystallization and data collection.** Before crystallization, NavAb was concentrated to ~20 mg ml<sup>-1</sup> and reconstituted into DMPC:CHAPSO (Anatrace) bicelles according to standard protocols<sup>54,55</sup>. The NavAb-bicelle preparation was mixed in a 1:1 ratio and setup in a hanging-drop vapour-diffusion format over a well solution containing 1.8–2.1 M ammonium sulphate, 100 mM Na-citrate pH 4.75. The mercury-free proteins, the mercury complexes, and the SeMet-labelled proteins all crystallized under essentially identical conditions. Crystals were typically passed through solutions containing 2 M ammonium sulphate, 100 mM Na-citrate pH 4.75 and 28% glucose (wt/v) in increments of ~6% glucose during harvesting. Crystals were plunged into liquid nitrogen and maintained at 100 K during all data collection procedures.

Over 1,000 crystals were screened and nearly 100 diffraction data sets were collected at a synchrotron radiation source (Advanced Light Source, BL8.2.1 and BL8.2.2). A SAD data set collected near the mercury absorption edge ( $\lambda = 1.005$  Å) from a mercury-containing complex of the NavAb(Ile217Cys) mutant was ultimately used to determine initial experimental phases. Our highest resolution SeMet SAD data set was collected near the selenium absorption edge ( $\lambda = 0.9795$  Å) from a mercury-free NavAb(Met221Cys) SeMet-labelled crystal. Subsequent native (that is, mercury-free) data sets were collected at standard wavelengths. Because the NavAb crystals were small (typically <0.15 mm × 0.15 mm × 0.15 mm), contained a high solvent content (~80%), were weakly diffracting, and radiation sensitive, special care was taken to minimize exposure times and to orient the crystals in order to maximize data completeness and quality.

**Structure determination and refinement.** X-ray diffraction data were integrated and scaled with the HKL2000 suite or DENZO/SCALEPACK<sup>56</sup> and, when required, further processed with the CCP4 package<sup>57</sup>. Experimental phases were determined using a 3.4 Å SAD data set from a Hg-containing NavAb(Ile217Cys) crystal. The SOLVE/RESOLVE<sup>58</sup> software were run in a standard setting and the first map, calculated at 3.7 Å, is shown in Supplementary Fig. 3. Ideal poly-alanine  $\alpha$ -helices were manually fitted into this map and the model was subsequently used in combined SAD-molecular replacement (MR) protocols within the Phenix software<sup>59</sup> using a 3.3 Å SAD data set obtained from a SeMet-labelled NavAb(Met221Cys) crystal. SAD-MR and MR-SAD-based maps were calculated and compared, allowing for complete register and amino acid assignment of the NavAb model. Higher-resolution native

data sets were ultimately obtained and phased by MR methods using the CNS suite<sup>60</sup> (although our best native NavAb(Met221Cys) data set is actually from a SeMet-containing crystal). Reiterative rounds of model building in O<sup>61</sup> were guided by inspection of omit maps and refinement with CNS<sup>60</sup> was performed with strict NCS-restraints, which were later relaxed during final rounds of refinement. Two strong densities (one per protein chain) assigned as solvent molecules (near the pore turret loop; not discussed in the main text) and all lipid molecules were added to the models at very late stages of refinement. Although trace amounts of digitonin are present in the crystallization condition, digitonin molecules were not readily observed in any electron density map. Refinement statistics, scaling statistics, and overall map quality were ultimately used to assign the NavAb space group as I222, although the data were found to closely mimic I422 ( $R_{\text{work}}/R_{\text{free}}$  stall at ~32% in I422).

**Structure analysis.** The geometry of NavAb structural models was assessed using PROCHECK<sup>62</sup>. The pore radius of NavAb was calculated using standard settings in the MOLE software<sup>63</sup>. Electrostatic surface calculations were performed with the APBS software<sup>64</sup>, calculated with 150 mM NaCl in the solvent. Structural alignments were performed using LSQMAN<sup>65</sup> and O<sup>61</sup>, where all channels were independently aligned onto NavAb based on the amino acid positions at the very beginning (that is, N-terminal portion) of their P-helices. The superposition of the atomic resolution Na<sup>+</sup>-complex structure<sup>40</sup> shown in Supplementary Fig. 12 was positioned manually, but the K<sup>+</sup>-channel and NaK-channel superpositions (Figs 2, 3, 5b and Supplementary Figs 12 and 13b) were obtained by simply aligning P-helices, as described earlier. All  $F_o - F_c$  omit maps shown throughout the main text and Supplementary Information have been calculated using standard settings and appropriate buffers in the CNS program<sup>60</sup>. The  $F_o - F_c$  omit map shown in Fig. 3b specifically derives from the 2.7 Å NavAb(Ile217Cys) data set and amino acids 170–183 were omitted from the calculation box. All structural figures were prepared with the PyMol software<sup>66</sup>.

**Electrophysiology.** NavAb was cloned into the CDM8 vector and transfected into tsA-201 cells (along with a CD8 marker construct) using standard protocols. Whole-cell currents were recorded with continuous perfusion of extracellular solution using an Axopatch 200 amplifier (Molecular Devices) with glass pipettes polished to 2–4 MΩ resistance. The intracellular pipette solution contained (in mM): 10 NaCl, 105 CsF, 20 TEA, 10 EGTA, 10 HEPES pH 7.4 (adjusted with CsOH). The extracellular Na<sup>+</sup> solution contained (in mM): 100 NaCl, 1 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 1 KCl, 50 TEA, 10 HEPES pH 7.4 (CsOH). For K-containing and Cs-containing extracellular solutions, NaCl was replaced with KCl or CsCl, respectively. The extracellular NMDG solution contained (in mM): 100 NMDG, 1 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 1 KCl, 50 TEA, 10 HEPES pH 7.4 (HCl) and the extracellular Ca<sup>2+</sup> solution contained (in mM): 75 CaCl<sub>2</sub>, 1 MgCl<sub>2</sub>, 1 KCl, 50 TEA, 10 HEPES pH 7.4 (CsOH). Voltage clamp pulses were generated and currents were recorded using Pulse software controlling an Instrutech ITC18 interface (HEKA). Data were analysed using Igor Pro 6.2 (WaveMetrics).

51. Koth, C. M. & Payandeh, J. Strategies for the cloning and expression of membrane proteins. *Adv. Protein Chem. Struct. Biol.* **76**, 43–86 (2009).
52. Cronin, C. N., Lim, K. B. & Rogers, J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci.* **16**, 2023–2029 (2007).
53. Chaptal, V. et al. Fluorescence Detection of Heavy Atom Labeling (FD-HAL): a rapid method for identifying covalently modified cysteine residues by phasing atoms. *J. Struct. Biol.* **171**, 82–87 (2010).
54. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316**, 1–6 (2002).
55. Faham, S. et al. Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Sci.* **14**, 836–840 (2005).
56. Otwinowski, Z. & Minor, W. *Processing of X-ray Diffraction Data Collected in Oscillation Mode* Vol. 276 (Academic, 1997).
57. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
58. Terwilliger, T. SOLVE and RESOLVE: automated structure solution and density modification. *Meth. Enzymol.* **374**, 22–37 (2003).
59. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
60. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
61. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
62. Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067 (1993).
63. Petřek, M., Kosinova, P., Koca, J. & Otyepka, M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* **15**, 1357–1363 (2007).
64. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
65. Kleywegt, G. J. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D* **52**, 842–857 (1996).
66. DeLano, W. L. PyMOL molecular viewer (v. 1.2r3pre) (<http://www.pymol.org>) (2002).



# DMRT1 prevents female reprogramming in the postnatal mammalian testis

Clinton K. Matson<sup>1,2</sup>, Mark W. Murphy<sup>1</sup>, Aaron L. Sarver<sup>3</sup>, Michael D. Griswold<sup>4</sup>, Vivian J. Bardwell<sup>1,2,3</sup> & David Zarkower<sup>1,2,3</sup>

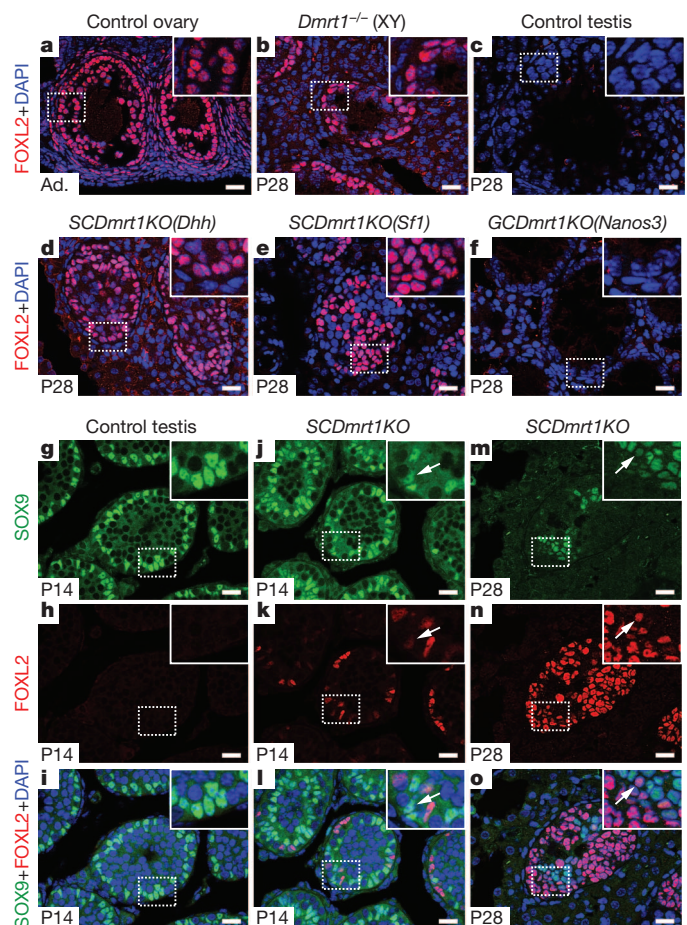
Sex in mammals is determined in the fetal gonad by the presence or absence of the Y chromosome gene *Sry*, which controls whether bipotential precursor cells differentiate into testicular Sertoli cells or ovarian granulosa cells<sup>1</sup>. This pivotal decision in a single gonadal cell type ultimately controls sexual differentiation throughout the body. Sex determination can be viewed as a battle for primacy in the fetal gonad between a male regulatory gene network in which *Sry* activates *Sox9* and a female network involving WNT/ $\beta$ -catenin signalling<sup>2</sup>. In females the primary sex-determining decision is not final: loss of the FOXL2 transcription factor in adult granulosa cells can reprogram granulosa cells into Sertoli cells<sup>2</sup>. Here we show that sexual fate is also surprisingly labile in the testis: loss of the DMRT1 transcription factor<sup>3</sup> in mouse Sertoli cells, even in adults, activates *Foxl2* and reprograms Sertoli cells into granulosa cells. In this environment, theca cells form, oestrogen is produced and germ cells appear feminized. Thus *Dmrt1* is essential to maintain mammalian testis determination, and competing regulatory networks maintain gonadal sex long after the fetal choice between male and female. *Dmrt1* and *Foxl2* are conserved throughout vertebrates<sup>4,5</sup> and *Dmrt1*-related sexual regulators are conserved throughout metazoans<sup>3</sup>. Antagonism between *Dmrt1* and *Foxl2* for control of gonadal sex may therefore extend beyond mammals. Reprogramming due to loss of *Dmrt1* also may help explain the aetiology of human syndromes linked to *DMRT1*, including disorders of sexual differentiation<sup>6</sup> and testicular cancer<sup>7</sup>.

Human chromosome 9p deletions removing *DMRT1* are associated with XY male-to-female sex reversal, and *Dmrt1* homologues determine sex in several non-mammalian vertebrates<sup>8–10</sup>. In mice, *Dmrt1* is expressed and required in both germ cells and Sertoli cells of the testis<sup>11–13</sup>. XY *Dmrt1*-null mutant mice are born as males with testes, although these gonads later undergo abnormal differentiation<sup>14</sup>; hence the role of *Dmrt1* in mammalian sex determination has been unclear (for overview of mammalian sex determination see Supplementary Fig. 1). Here we examine *Dmrt1* mutant testes during postnatal development, asking whether loss of *Dmrt1* causes postnatal feminization in mice.

We first examined gonads of *Dmrt1*-null mutant males (*Dmrt1*<sup>−/−</sup>) for the presence of FOXL2, a female-specific transcription factor expressed in granulosa cells and theca cells<sup>15,16</sup>, the two somatic cell types of the ovarian follicle (Fig. 1a). Four weeks after birth, abundant FOXL2-positive cells were present within mutant seminiferous tubules (Fig. 1b), which in control testes contain only germ cells and Sertoli cells (Fig. 1c). To establish the origin of the FOXL2-positive cells, we deleted *Dmrt1* either in germ cells (using *Nanos3-cre*) or in Sertoli cells (using *Dhh-cre* or *Sf1-cre*) (Supplementary Fig. 2a–l and Supplementary Table 1). Loss of *Dmrt1* in fetal Sertoli cells (*SCDmrt1KO*) but not in fetal germ cells (*GCDmrt1KO*) induced FOXL2 expression (Fig. 1d–f). *SCDmrt1KO* gonads retained small numbers of germ cells, which appeared to arrest in meiotic prophase on the basis of SYCP3 localization (Supplementary Fig. 3). These results demonstrate that DMRT1

expression in Sertoli cells prevents FOXL2 expression and suggest that *Dmrt1* mutant testes become feminized during the first postnatal month.

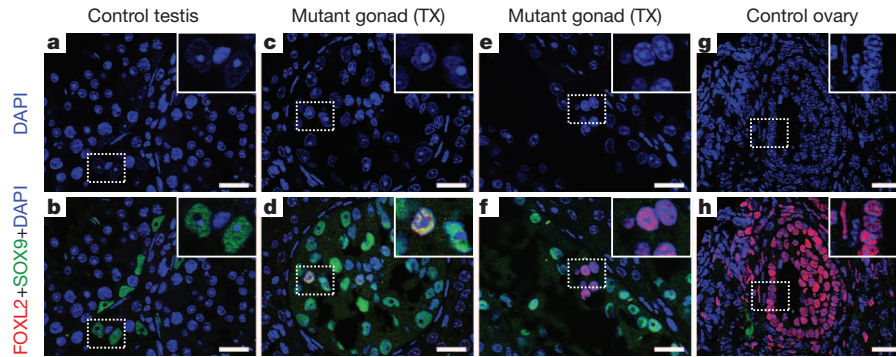
Next we examined the timing of FOXL2 induction. At postnatal day (P)7, *SCDmrt1KO* testes had seminiferous tubules in which all Sertoli cells expressed SOX9 normally (Supplementary Fig. 2m–r), but at P14



**Figure 1 | DMRT1 maintains SOX9 and suppresses FOXL2 expression in postnatal Sertoli cells.** a–c, FOXL2 expression detected by immunofluorescence in adult (Ad.) granulosa and theca cells of control ovary (a) and intratubular cells of *Dmrt1*-null testis at P28 (b), but not in control testis (c). DAPI, 4',6-diamidino-2-phenylindole. d–f, FOXL2 is robustly expressed when *Dmrt1* is mutated in fetal Sertoli cells with *Dhh-cre* (d) or *Sf1-cre* (e) but not when *Dmrt1* is mutated in fetal germ cells with *Nanos3-cre* (f). g–o, Timing of FOXL2 expression. FOXL2 is absent from control testis at P14 (g–i). Cells expressing FOXL2 or FOXL2 and SOX9 (arrowheads) are present in *SCDmrt1KO* testis at P14 (j–l). FOXL2-positive cells are abundant in *SCDmrt1KO* testis at P28 and most cells no longer express SOX9 (m–o). Scale bars, 20  $\mu$ m.

<sup>1</sup>Developmental Biology Center and Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis 55455, Minnesota, USA. <sup>2</sup>Molecular, Cellular, Developmental Biology, and Genetics Graduate Program, University of Minnesota, Minneapolis 55455, Minnesota, USA. <sup>3</sup>University of Minnesota Masonic Cancer Center, Minneapolis, Minnesota 55455, USA. <sup>4</sup>School of Molecular Biosciences, Washington State University, Pullman, Washington 99164, USA.

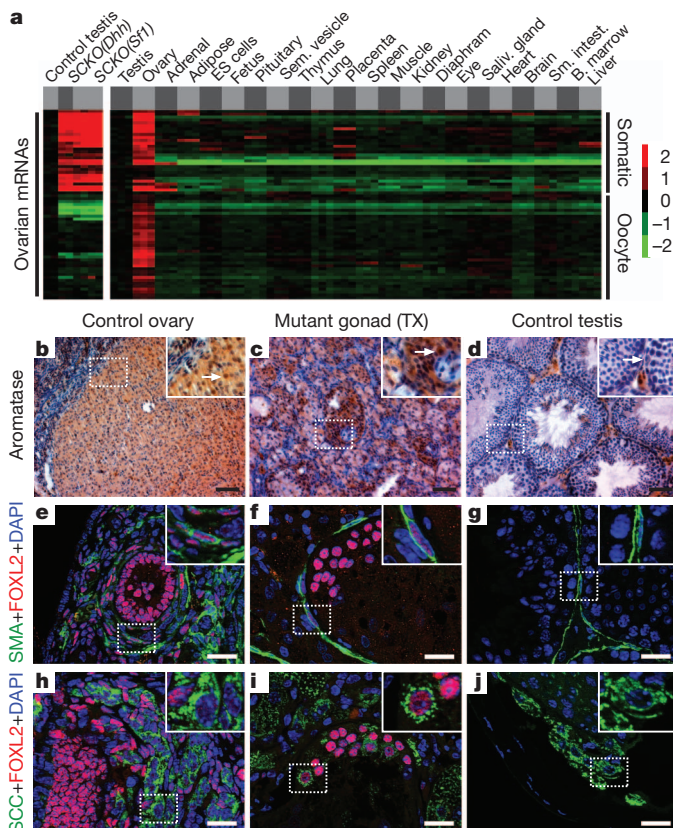




**Figure 2 | Sertoli-to-granulosa transdifferentiation in the adult testis.** **a–h**, Expression of FOXL2 and SOX9 one month after tamoxifen (TX) injection into *Dmrt1<sup>lox/lox</sup>* adult males (8 weeks and older) carrying inducible ubiquitous *cre* transgene *UBC-cre/ERT2*. **a, b**, Sertoli cells in control testis express SOX9 but not FOXL2. **c–f**, Mutant testis has Sertoli-like cells expressing SOX9 or SOX9 and FOXL2 (**d**, inset) and granulosa-like cells expressing only FOXL2

some intratubular cells co-expressed SOX9 and FOXL2 or lacked SOX9 and strongly expressed FOXL2 (Fig. 1g–l). By P28 few SOX9-positive cells remained and most intratubular cells strongly expressed FOXL2 (Fig. 1m–o). Histological analysis of mutant gonads is shown in Supplementary Fig. 4. These results show that fetal loss of *Dmrt1* causes postnatal Sertoli cells to lose the male-promoting SOX9 and instead express the female-promoting FOXL2.

Loss of *Foxl2* in the adult ovary can lead to transdifferentiation of granulosa cells to Sertoli cells<sup>2</sup>, so we asked whether loss of *Dmrt1* in the adult testis activates *Foxl2* and causes the reciprocal sex transformation, from Sertoli to granulosa. Indeed, one month after deletion of *Dmrt1* in adult males (using a tamoxifen-inducible *cre* transgene), we observed cells with typical Sertoli cell features including tripartite nucleoli but expressing both SOX9 and FOXL2 (Fig. 2a–d), as well as



**Figure 3 | Feminization of *SCDmrt1KO* XY gonads.** **a**, Expression of ovary-enriched mRNAs with expression profiles similar to *Foxl2* (see Supplementary Information). mRNAs labelled 'somatic' were enriched in ovarian somatic cells; those labelled 'oocyte' were enriched in female germ cells. See Supplementary Fig. 6 for higher resolution image. **b**, marrow, bone marrow; Saliv. gland, salivary gland; Sem. vesicle, seminal vesicle; Sm. intestine, small intestine. **b–d**, Immunohistochemistry detection of CYP11A1/aromatase expression in follicles of control adult ovary (**a**) and in adult XY *SCDmrt1KO* gonad (**b**), but only in interstitial Leydig cells of control testis (**c**). Arrows indicate aromatase-positive granulosa cells in ovary and mutant gonad and negative Sertoli cell in control testis. Scale bars, 50  $\mu$ m. **e–g**, Immunofluorescence detection of SMA and FOXL2. Ovarian theca cells (**e**, inset) are elongated cells expressing both proteins; similar cells are present in mutant gonads (**f**); peritubular myoid cells in control testes express SMA and not FOXL2 (**g**). Scale bars, 20  $\mu$ m. **h–j**, Immunofluorescence detection of cells coexpressing FOXL2 in the nucleus and steroidogenic enzyme CYP11A1/SCC at high levels in the cytoplasm in control ovary (**h**) and XY *Dmrt1KO* gonads (**i**). SCC-positive cells in control testis (**j**) are interstitial Leydig cells. Mutant gonads were *SCDmrt1KO(Dhh)*. All tissues were from adult mice. Scale bars, 20  $\mu$ m.

(**f**, inset). **g, h**, FOXL2-positive cells in control ovary have DAPI morphology similar to FOXL2 single-positive cells of mutant testis. FOXL2-positive cells in mutant testis resemble granulosa cells: they lack the tripartite nucleoli of Sertoli cells, have smaller and more rounded nuclei, and have more punctate DAPI staining. *UBC-cre/ERT2* also deletes *Dmrt1* in germ cells, causing precocious meiosis<sup>12</sup>; after one month germ cells are nearly absent. Scale bars, 20  $\mu$ m.

cells with typical granulosa cell nuclear morphology that lacked SOX9 and strongly expressed FOXL2 (Fig. 2e–h). Thus antagonism between DMRT1 and FOXL2 continues into adulthood and Sertoli cell fate remains plastic even after terminal differentiation.

To evaluate further the transformation of mutant gonads, we compared the messenger RNA profile of control and mutant P28 testes; 5,030 mRNAs were expressed >8-fold differently across this data set or a data set comparing testis to 21 other tissues including ovary (Supplementary Fig. 5a). We calculated Pearson correlation coefficients for expression of these 5,030 mRNAs in mutant gonads relative to each tissue and found that the mutant gonad most closely resembled ovary (Supplementary Fig. 5b; average  $R = 0.75$ ). Many mRNAs with decreased expression in mutant gonads also were low in other tissues, probably reflecting a lack of male germ cells, which comprise much of the testis mass. Also, some mRNAs that were increased in mutant gonads were increased in other tissues. Therefore, to specifically evaluate ovary-enriched mRNAs, we used bioGPS (<http://www.biogps.gnf.org>; see Supplementary Information) to identify 65 mRNAs with expression closely correlated to *Foxl2* and then compared their expression in ovary relative to the other 21 tissues (Fig. 3a and Supplementary Fig. 6). This comparison confirmed that these mRNAs are highly ovary enriched. About 40% were increased in mutant gonads relative to control testes; about 80% of the remainder were oocyte enriched. Thus loss of *Dmrt1* causes large changes in mRNA expression, including induction of multiple ovary-enriched mRNAs. mRNA profiling of *Dmrt1* mutant gonads perinatally and at P9 did not reveal

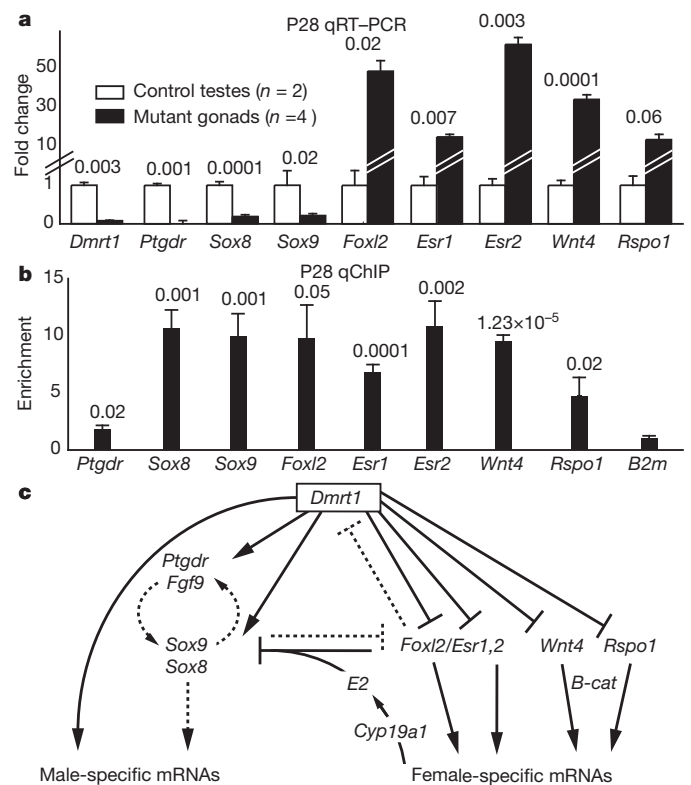
apparent feminization<sup>17,18</sup>, consistent with the observation that FOXL2 expression starts at ~P14.

Further analysis of the mRNA profiling data identified highly increased expression (>5-fold,  $P < 0.001$ ) of many mRNAs expressed in granulosa cells and required for ovarian development or function. These included *Foxl2*, *Nr5a2* (also known as *Lrh1*), *Wnt4*, LH receptor (*Lhcgr*), prolactin receptor (*Prlr*), FSH receptor (*Fshr*), follistatin (*Fst*), *Sfrp4*, *Igfbp5*, *Inhbb*, *Inha* and *Lnfg* (Supplementary Table 2). *Foxl2os*, a noncoding RNA transcribed from the opposite strand of the *Foxl2* coding region, also was highly overexpressed and has been suggested as a positive regulator of *Foxl2* (ref. 19). We confirmed increased expression in mutant gonads of LRH1, a transcription factor expressed only in granulosa cells within the ovary<sup>20</sup> and absent from the testis (Supplementary Fig. 7a–f). *Nr5a2* is probably a direct target of DMRT1 regulation, based on binding of DMRT1 to its promoter proximal sequences *in vivo* (Supplementary Fig. 7g). On the basis of mRNA and protein expression data and changes in cellular morphology, we conclude that loss of *Dmrt1* in testes reprograms Sertoli cells into granulosa cells.

Granulosa cells produce oestrogens, which are essential for ovarian development in many vertebrates; in mammals, oestrogen signalling also acts with FOXL2 to repress *Sox9* transcription in adult granulosa cells<sup>2</sup>. HSD17 $\beta$ 1 and CYP19A1/aromatase are enzymes critical for oestrogen synthesis, and mRNAs for both enzymes were increased in mutant gonads (Supplementary Fig. 8). Aromatase protein is robustly expressed in granulosa cells and was strongly expressed in mutant gonads (Fig. 3b–d). Consistent with these enzyme changes, oestradiol was raised in the serum of adult mutants relative to control adult males (Supplementary Information). Although expression of the androgenic enzyme *Hsd17 $\beta$ 3* was not affected in mutant gonads (Supplementary Fig. 8), androgen levels were reduced based on severely decreased seminal vesicle weight, a sensitive indicator of androgen activity ( $350 \pm 52$  mg versus  $182 \pm 36$  mg;  $n = 3$ ,  $P = 0.01$ ).

Theca cells are induced during follicle growth in the ovary, probably in response to granulosa cell signals<sup>21</sup>, and together with granulosa cells and oocytes they comprise the functional unit of the ovary. Because mutant gonads contained apparently functional granulosa cells, we asked whether theca cells also formed. Theca cells have spindle-shaped nuclei and express both FOXL2 and smooth muscle actin (SMA) (Fig. 3e). Adult mutant gonads contained cells closely resembling theca cells and expressing both proteins (Fig. 3f). The theca-like cells probably derive either from granulosa cells or peritubular myoid cells (which also are elongated and express SMA; Fig. 3g). However, as seminiferous tubule integrity was lost before formation of these cells (Fig. 3f and Supplementary Fig. 9), they could potentially derive from interstitial cells that invaded the tubule remnants. We also observed intratubular cells strongly expressing the steroidogenic enzyme SCC (Fig. 3h–j); these cells resembled luteinized granulosa cells of the ovary (Fig. 3h), suggesting that granulosa cells in the mutant gonad are responsive to gonadotropins. We therefore tested the effect of exogenous gonadotropin stimulation; treated mutants, but not controls, had additional luteinized granulosa cells and germ cells with oocyte-like nuclear morphology that expressed the oocyte-specific proteins MATER and ZP2 (Supplementary Fig. 10). This result indicates that both somatic cells and germ cells are feminized in mutant gonads.

The preceding results indicate that DMRT1 is essential for postnatal sex maintenance. DMRT1 is a sequence-specific transcriptional regulator, capable of activating or repressing transcription of target genes<sup>18,22</sup>. To help find targets of DMRT1 regulation with potential roles in sex maintenance we examined expression of known fetal sex-determining genes in mutant gonads at P28 by quantitative polymerase chain reaction with reverse transcription (qRT–PCR; Fig. 4a). Among masculinizing genes, *Ptgd*, *Sox9* and *Sox8*, which acts redundantly with *Sox9* (refs 23, 24), were reduced. Among feminizing genes, *Foxl2*, *Esr1*, *Esr2*, *Wnt4* and *Rspo1* were increased. We assayed binding of DMRT1 to DNA of P28 testes by quantitative chromatin immunoprecipitation



**Figure 4 | DMRT1 regulation of postnatal gene expression.** **a**, qRT–PCR analysis of sex-determining genes at P28. Significance of expression changes is indicated (Student's *t*-test). Mutant gonads were *SCDmrt1KO(Sf1)*; *SCDmrt1KO(Dhh)* mutant gonads and equivalent expression changes. **b**, qChIP analysis of DMRT1 DNA binding in P28 testes. Significance of enrichment relative to *B2m* (Student's *t*-test) is shown. **c**, Model for regulation by postnatal sex maintenance by DMRT1. Proposed direct regulation based on ChIP and mRNA expression data are indicated by solid lines; indirect or potential regulation is indicated by dashed lines. Model adapted from ref. 2.

(qChIP), guided by genome-wide ChIP data from P9 testes (ChIP-chip<sup>18</sup> and ChIP-seq (unpublished data)). DMRT1 bound both upstream and downstream of *Sox9* and upstream of *Sox8*, and bound weakly near *Ptgd*. DMRT1 bound strongly near *Foxl2*, *Esr1*, *Esr2*, *Wnt4* and *Rspo1* (Fig. 4b). All of the DMRT1-associated regions contained at least one close match to the DMRT1 DNA-binding consensus<sup>18,22</sup>.

On the basis of mRNA and protein expression data and ChIP analysis, we propose a model for postnatal sex maintenance (Fig. 4b) in which DMRT1 maintains male fates by repressing multiple female-promoting genes and activating male-promoting genes. *Sox9* is dispensable for testis differentiation after sex determination<sup>24,25</sup>, suggesting that other critical male regulators remain to be found; *Sox8* is a clear candidate based on its redundancy with *Sox9* (refs 23, 24). We find that DMRT1 represses *Foxl2*, which is known to maintain postnatal ovarian fate. FOXL2 also represses *Dmrt1* (ref. 2); thus antagonism between these sex-specific transcriptional regulators may be central to sex maintenance in both sexes throughout reproductive life. *Wnt4* and *Rspo1* also are prime candidates for postnatal sex maintenance based on their requirement in ovarian determination in the fetus<sup>26,27</sup>. Indeed, P28 mutant gonads had increased nuclear  $\beta$ -catenin in somatic cells, as in ovaries, but control testes did not, indicating active WNT/ $\beta$ -catenin signalling in the mutant gonads (Supplementary Fig. 11). Functional analysis of *Wnt4*, *Rspo1* and other known fetal sex regulators will be important to establish their roles in sex maintenance.

The analysis presented here demonstrates that deletion of *Dmrt1* during fetal development induces postnatal feminization of the testis, causing male-to-female primary sex reversal. Moreover, deletion of



*Dmrt1* in adults can reprogram differentiated Sertoli cells into apparent granulosa cells. Why *Dmrt1* mutants are feminized only after birth remains unclear. Another male-promoting gene may act redundantly with *Dmrt1* before P14, masking its function; alternatively, the testis may lack potential feminizing activity from genes such as *Foxl2* before P14. Another puzzle is that *Dmrt1* mutant mice are born male, whereas human 9p deletions removing *DMRT1* can cause XY feminization at birth. The human sex reversal may reflect failure to maintain male sex determination, and the longer human gestation may permit testis-to-ovary reprogramming before birth. Alternatively, human testes may have potential feminizing activity earlier or may lack masculinizing genes redundant with *DMRT1*. Our results may provide insights into the aetiology of human gonadal disorders, including gonadoblastoma and granulosa cell tumours of the testis. Moreover, because many genes implicated in this study are evolutionarily conserved, similar mechanisms may control adult sex switching in fish and may maintain sexual fate in the adult gonads of other vertebrates or even in other phyla.

## METHODS SUMMARY

**Mouse breeding.** *Dmrt1* mutant and control males were generated as described<sup>12</sup>; tissue-specific Cre recombinase strains are in Supplementary Table 1. Adult wild-type or *Dmrt1*<sup>flx/flx</sup> females were used as controls. Mice were mixed C57BL/6J, 129S1 and FVB genetic background. Protocols were approved by the Institutional Animal Care and Use Committee.

**Immunofluorescence and immunohistochemistry.** Immunofluorescence and immunohistochemistry were performed as described<sup>12</sup>. Antibodies are listed in Supplementary Table 3. Analyses included at least two biological replicates.

**Tamoxifen treatment.** Tamoxifen-inducible deletion of *Dmrt1* in adult males was as described<sup>12</sup>. Testes were harvested one to two months after treatment.

**mRNA expression analysis.** mRNA expression profiling and data analysis were as described<sup>13</sup> except total testis RNA was isolated from 4-week-old mice using TRIzol reagent (Invitrogen no. 15596-026). Additional detail is in Supplementary Methods.

**qRT-PCR.** qRT-PCR was as described<sup>12</sup>. qRT-PCR primers are listed in Supplementary Table 4.

**ChIP.** ChIP followed by either microarray (ChIP-chip) or qPCR analysis (qChIP) were as described<sup>18</sup>. Gene-specific primers used for qChIP are in Supplementary Table 4.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 February; accepted 27 May 2011.

Published online 20 July 2011.

- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117–121 (1991).
- Uhlenhaut, N. H. *et al.* Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**, 1130–1142 (2009).
- Raymond, C. S. *et al.* Evidence for evolutionary conservation of sex-determining genes. *Nature* **391**, 691–695 (1998).
- Loffler, K. A., Zarkower, D. & Koopman, P. Etiology of ovarian failure in blepharophimosis ptosis epicanthus inversus syndrome: *FOXL2* is a conserved, early-acting gene in vertebrate ovarian development. *Endocrinology* **144**, 3237–3243 (2003).
- Raymond, C. S., Kettlewell, J. R., Hirsch, B., Bardwell, V. J. & Zarkower, D. Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev. Biol.* **215**, 208–220 (1999).
- Tannour-Louet, M. *et al.* Identification of *de novo* copy number variants associated with human disorders of sexual development. *PLoS ONE* **5**, e15392 (2010).
- Turnbull, C. *et al.* Variants near *DMRT1*, *TERT* and *ATF7IP* are associated with testicular germ cell cancer. *Nature Genet.* **42**, 604–607 (2010).

- Yoshimoto, S. *et al.* A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc. Natl Acad. Sci. USA* **105**, 2469–2474 (2008).
- Smith, C. A. *et al.* The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).
- Matsuda, M. *et al.* *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417**, 559–563 (2002).
- Kim, S., Bardwell, V. J. & Zarkower, D. Cell type-autonomous and non-autonomous requirements for *Dmrt1* in postnatal testis differentiation. *Dev. Biol.* **307**, 314–327 (2007).
- Matson, C. K. *et al.* The mammalian doublesex homolog *DMRT1* is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Dev. Cell* **19**, 612–624 (2010).
- Krentz, A. D. *et al.* The DM domain protein *DMRT1* is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. *Proc. Natl Acad. Sci. USA* **106**, 22323–22328 (2009).
- Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J. & Zarkower, D. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* **14**, 2587–2595 (2000).
- Schmidt, D. *et al.* The murine winged-helix transcription factor *Foxl2* is required for granulosa cell differentiation and ovary maintenance. *Development* **131**, 933–942 (2004).
- Uda, M. *et al.* *Foxl2* disruption causes mouse ovarian failure by pervasive blockage of follicle development. *Hum. Mol. Genet.* **13**, 1171–1181 (2004).
- Fahrioglu, U., Murphy, M. W., Zarkower, D. & Bardwell, V. J. mRNA expression analysis and the molecular basis of neonatal testis defects in *Dmrt1* mutant mice. *Sex. Dev.* **1**, 42–58 (2007).
- Murphy, M. W. *et al.* Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog *DMRT1* in the juvenile testis. *Proc. Natl Acad. Sci. USA* **107**, 13360–13365 (2010).
- Cocquet, J., Pannetier, M., Fellous, M. & Veitia, R. A. Sense and antisense *Foxl2* transcripts in mouse. *Genomics* **85**, 531–541 (2005).
- Duggavathi, R. *et al.* Liver receptor homolog 1 is essential for ovulation. *Genes Dev.* **22**, 1871–1876 (2008).
- Orisaka, M., Tajima, K., Tsang, B. K. & Kotsuji, F. Oocyte-granulosa-theca cell interactions during preantral follicular development. *J. Ovarian Res.* **2**, 9 (2009).
- Murphy, M. W., Zarkower, D. & Bardwell, V. J. Vertebrate DM domain proteins bind similar DNA sequences and can heterodimerize on DNA. *BMC Mol. Biol.* **8**, 58 (2007).
- Chaboissier, M. C. *et al.* Functional analysis of *Sox8* and *Sox9* during sex determination in the mouse. *Development* **131**, 1891–1901 (2004).
- Barrionuevo, F. *et al.* Testis cord differentiation after the sex determination stage is independent of *Sox9* but fails in the combined absence of *Sox9* and *Sox8*. *Dev. Biol.* **327**, 301–312 (2009).
- Chang, H. *et al.* *Wt1* negatively regulates  $\beta$ -catenin signaling during testis development. *Development* **135**, 1875–1885 (2008).
- Vainio, S., Heikkilä, M., Kispert, A., Chin, N. & McMahon, A. P. Female development in mammals is regulated by *Wnt-4* signalling. *Nature* **397**, 405–409 (1999).
- Parma, P. *et al.* *R-spondin1* is essential in sex determination, skin differentiation and malignancy. *Nature Genet.* **38**, 1304–1309 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Treier for helpful discussion, K. Hatzl, A. Minkina, A. Peterson, the University of Minnesota Mouse Genetics Laboratory and C. Small for technical assistance, J. Dean, R. Veitia and K.-i. Morohashi for antibodies, D. Greenstein and A. M. Weber-Main for comments on the manuscript, C. Manivel for histology expertise, M. Steffes and D. Gabrielson for oestradiol analysis, and the University of Minnesota Supercomputing Institute for computational resources. This work was funded by the NIH (GM59152), the Minnesota Medical Foundation, and a predoctoral fellowship from the NSF (to C.K.M.).

**Author Contributions** C.K.M. performed mouse breeding and analysis of protein and mRNA expression; M.W.M. performed ChIP analysis; A.L.S. performed bioinformatic analysis; C.K.M., D.Z. and V.J.B. designed the study, analysed data, and wrote the paper; M.D.G. provided mRNA profiling expertise; all authors discussed the results and edited the paper.

**Author Information** mRNA expression profiling data have been deposited at the Gene Expression Omnibus under accession number GSE27261 and can be reviewed at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=z1ctbiugsamymtc&acc=GSE27261>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.Z. (zarko001@umn.edu) or V.J.B. (bardw001@umn.edu).



## METHODS

**Mouse breeding.** Conditional *Dmrt1* mutant and control males were generated as described<sup>12</sup>; tissue-specific Cre recombinase strains are in Supplementary Table 1. Adult wild-type or *Dmrt1*<sup>fllox/fllox</sup> females were used as controls. Mice were of mixed C57BL/6J, 129S1 and FVB genetic background. Protocols were approved by the University of Minnesota Institutional Animal Care and Use Committee.

**Immunofluorescence and immunohistochemistry.** Both immunofluorescence and immunohistochemistry were performed as described<sup>12</sup>. Antibodies are listed in Supplementary Table 3. Analyses included a minimum of two biological replicates.

**Tamoxifen treatment.** Tamoxifen-inducible deletion of *Dmrt1* in adult males was performed as previously described<sup>12</sup>. Testes were harvested one to two months after treatment.

**mRNA expression analysis.** mRNA expression profiling and data analysis were performed as described<sup>13</sup> except total testis RNA was isolated from 4-week-old mice using TRIzol reagent (Invitrogen no. 15596-026). Affymetrix Mouse Genome 439 2.0 arrays were normalized by GC-RMA normalization<sup>28</sup> using GeneData Refiner. The Raw .cel files and the normalized data are deposited in the Gene Expression Omnibus (GEO)<sup>29</sup> under accession number GSE27261. GSE9954 was obtained from the GEO database. The arrays with the highest sample identification numbers were removed from the tissue data set to select 22 tissue types, each with three experimental replicates. When multiple probe sets were mapped to the same gene symbol, these values were averaged to obtain one value for each gene symbol. Direct Pearson correlation *R* values were calculated using all array data following reduction to gene symbols, and these values are shown in Fig. 2b.

Each experiment in our data set was divided by the average expression value from control testis tissue. GSE9954 data were separately divided by the average signal obtained from the GSE9954 testis samples. This was done separately for each data set to determine how samples from each data set differed from a baseline 'testis' expression state. Cluster 3.0 software<sup>30</sup> was used to: (1) log base 2 transform

the data; (2) filter the data set for genes that showed at least three observations with  $\text{abs}(\text{val}) > 3$  (eightfold), which resulted in 5,030 genes passing the filter using both data sets combined; and (3) cluster the data on the gene axis using average linkage hierarchical clustering. The experimental axis was defined by order of decreasing correlation to the mutant testes calculated as described earlier. Javatreview Software<sup>31</sup> was used to generate heatmap images.

**qRT-PCR.** qRT-PCR was performed as described<sup>12</sup>. qRT-PCR primers are listed in Supplementary Table 4.

**Chromatin immunoprecipitation.** ChIP followed by either microarray (ChIP-chip) or qPCR analysis (qChIP) were performed as described<sup>18</sup>. Gene-specific primers used for qChIP are in Supplementary Table 4.

**Oestradiol assays.** Serum oestradiol was assayed using a clinical electrochemiluminescence immunoassay (Roche Estradiol II, 03000079 122) according to manufacturer's instructions. Three of three males assayed had levels below the detection limit, whereas two of three females had measurable oestradiol (5.0 and 19.7 pg dl<sup>-1</sup>). Two of three *SCDmrt1KO(Dhh)* mutant males had measurable oestradiol (5.6 and 21.2 pg dl<sup>-1</sup>).

**Gonadotropin treatment.** Six-to-eight-week-old mutant males, control males and control females were treated with 5 units of pregnant mare serum by intraperitoneal injection and gonads were harvested 48 h later.

28. Wu, J., Irazarray, R. A., Gentleman, R., Martinez-Murillo, F. & Spencer, F. A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.* **99**, 909–917 (2004).
29. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
30. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
31. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).

# A two-step chemical mechanism for ribosome-catalysed peptide bond formation

David A. Hiller<sup>1</sup>, Vipender Singh<sup>1†</sup>, Minghong Zhong<sup>1†</sup> & Scott A. Strobel<sup>1</sup>

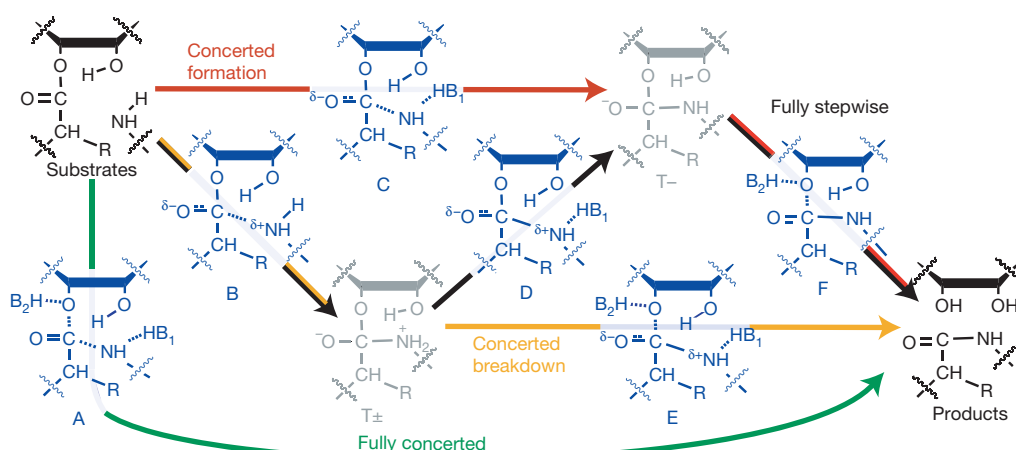
The chemical step of natural protein synthesis, peptide bond formation, is catalysed by the large subunit of the ribosome. Crystal structures have shown that the active site for peptide bond formation is composed entirely of RNA<sup>1</sup>. Recent work has focused on how an RNA active site is able to catalyse this fundamental biological reaction at a suitable rate for protein synthesis. On the basis of the absence of important ribosomal functional groups<sup>2</sup>, lack of a dependence on pH<sup>3</sup>, and the dominant contribution of entropy to catalysis<sup>4</sup>, it has been suggested that the role of the ribosome is limited to bringing the substrates into close proximity. Alternatively, the importance of the 2'-hydroxyl of the peptidyl-transfer RNA<sup>5</sup> and a Brønsted coefficient near zero<sup>6</sup> have been taken as evidence that the ribosome coordinates a proton-transfer network. Here we report the transition state of peptide bond formation, based on analysis of the kinetic isotope effect at five positions within the reaction centre of a peptidyl-transfer RNA mimic. Our results indicate that in contrast to the uncatalysed reaction, formation of the tetrahedral intermediate and proton transfer from the nucleophilic nitrogen both occur in the rate-limiting step. Unlike in previous proposals, the reaction is not fully concerted; instead, breakdown of the tetrahedral intermediate occurs in a separate fast step. This suggests that in addition to substrate positioning, the ribosome is contributing to chemical catalysis by changing the rate-limiting transition state.

Several reaction mechanisms have been proposed for peptide bond formation (Fig. 1). The peptidyl transferase reaction occurs through nucleophilic attack of the  $\alpha$ -amino group of aminoacyl-tRNA on the carbonyl carbon of the peptidyl-tRNA. A peptide bond forms and the

ester bond linking the peptide to the 3'-oxygen of peptidyl-tRNA breaks, leaving a deacylated tRNA and a peptide lengthened by one amino acid. If substrate positioning is the sole contribution to catalysis, the mechanism within the ribosome is expected to be equivalent to the well-studied uncatalysed reactions, where a pathway involving two tetrahedral intermediates is followed ( $T^{\pm}$  and  $T^{-}$ , Fig. 1; black pathway, see figure legend for explanation of mechanisms and intermediates). At high pH, the rate-limiting transition state for the uncatalysed reaction was predicted<sup>7</sup> to occur during deprotonation of the zwitterionic  $T^{\pm}$  intermediate, on the basis of the pH-rate dependence of the reaction and a Brønsted coefficient near 1 (transition state D in Fig. 1). At low pH, breakdown of the negatively charged  $T^{-}$  is rate-limiting (transition state F in Fig. 1).

It has also been suggested<sup>8,9</sup> that the peptidyl-tRNA 2'-hydroxyl acts as a 'proton shuttle' that abstracts a proton from the nucleophilic amine and simultaneously donates another proton either directly or indirectly to the leaving group oxygen. This could happen in a stepwise fashion (Fig. 1, red and orange pathways), involving a tetrahedral intermediate. Alternatively, the reaction has been proposed to be fully concerted, with the amide bond to the nucleophile formed at the same time that the ester bond to the leaving group is broken (Fig. 1, green pathway). Recently, computational methods were used to evaluate concerted proton shuttle mechanisms with and without an additional water molecule, and favoured a fully concerted eight-member proton shuttle (similar to transition state A)<sup>10</sup>. Alternatively, other computational studies have indicated a stepwise proton shuttle would be favourable<sup>11,12</sup>. All of these possibilities have been proposed<sup>4,8–13</sup>.

Each mechanism predicts a different transition state and hence a different role for the ribosome in catalysis. Therefore we sought to



**Figure 1 | Proposed reaction mechanisms.** Ground states are shown in black, intermediates in grey and transition states in blue. The ribosomal reaction may be fully stepwise with two intermediates, like the uncatalysed reaction (black pathway). Nucleophilic attack leads to an intermediate with a positively charged nitrogen and negatively charged carbonyl oxygen ( $T^{\pm}$ ).

Deprotonation leads to a negatively charged intermediate ( $T^{-}$ ), which breaks down to products. Alternatively, one of these steps may be concerted, with only one intermediate (red and orange pathways) or the reaction may be fully concerted with no intermediates (green). The identities of B1 and B2 are uncertain; the 2'-hydroxyl may be one, both or neither.

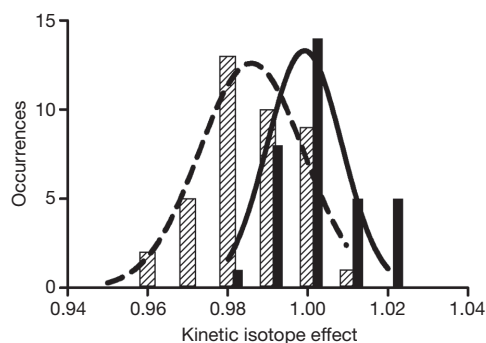
<sup>1</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. <sup>†</sup>Present addresses: Department of Chemistry and Bioengineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (V.S.); AMRI, 30 Corporate Circle, Albany, New York 12203, USA (M.Z.).

obtain experimental constraints for modelling the transition state of ribosome-catalysed peptide bond formation using kinetic isotope effect (KIE) analysis (reviewed in ref. 14). Generally, if the rate effect of isotopic substitution (defined as  $k_{\text{light}}/k_{\text{heavy}}$ ) is greater than one (termed 'normal'), it indicates that bonding to that isotope is weaker in the transition state. Reaction coordinate motion of a substituted atom will also contribute to a normal isotope effect. Alternatively, if the isotope effect is less than one (termed 'inverse'), it indicates stronger bonding. The magnitude of the effect is correlated with the change in bond order. To make this analysis possible, we previously synthesized a complete series of molecules that differ by a single isotopic substitution at each atom within the reaction centre<sup>15,16</sup>. These molecules are substrates for the ribosomal 50S fragment reaction<sup>17,18</sup>. Previous studies indicate that chemistry is rate-limiting for this assay and that unlike the 70S reaction, which has a complete commitment to catalysis, this reaction is amenable to KIE analysis<sup>18</sup>. The Brønsted coefficient of the nucleophile and the structure of the active site are within experimental error for 50S and 70S ribosomes, indicating that the mechanism of catalysis is similar<sup>6,19</sup>.

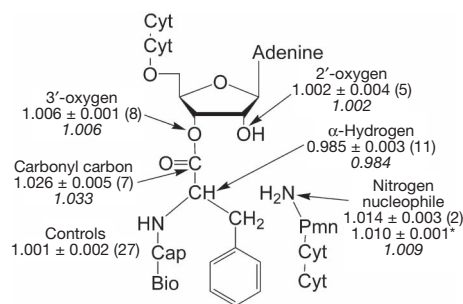
To achieve the necessary precision, KIEs were measured by a competitive assay. The light and heavy substrates were incubated in the same reaction, and the change in their ratio as the reaction proceeded was used to determine their relative reaction rates. We used two remote radiolabels, <sup>32</sup>P and <sup>33</sup>P, attached to the 5'-ends of the P-site substrate and performed scintillation counting to define the relative abundance of the two substrates, as previously described for the uncatalysed reaction<sup>20,21</sup>.

Given the small magnitude of the expected effects (typically in the range of 0.95 to 1.05) and the complexity of the experimental system (a 1.5-MDa complex altered by 1–2 Da), we performed several controls to determine the extent of random and systematic error. First, we demonstrated that the effect from the remote label alone ( $1.001 \pm 0.002$ , mean  $\pm$  s.e.m.) is much smaller than the effects we expect for atoms at the reaction centre. This value and the effect of deuterium substitution at the amino acid  $\alpha$ -carbon were determined 33 and 40 times, respectively. Repeated measurements in both cases yielded normally distributed data, with a significant difference between the mean of the control and the mean of the  $\alpha$ -deuterium substitution (Fig. 2). Furthermore, two sets of measurements were made for each substitution, one pairing <sup>32</sup>P with the light isotope at the reaction centre and <sup>33</sup>P with the heavy isotope, and a second with the opposite pairings—that is, <sup>32</sup>P with the heavy isotope and <sup>33</sup>P with the light isotope. These determinations gave the same values within experimental error.

We measured isotope effects for five positions at the reaction centre: the 3'-oxygen leaving group, the carbonyl carbon, the nucleophilic nitrogen, the hydrogen attached to the  $\alpha$ -carbon, and the vicinal 2'-hydroxyl (Fig. 3). The simplest effect to interpret is that of <sup>18</sup>O substitution on the leaving group, which measures the extent of C–O bond dissociation in the transition state. Effects in the range of



**Figure 2 | Histograms of measured isotope effects.** Isotope effects on peptide bond formation for controls (solid bars, solid line) and  $\alpha$ -deuterium substitution (hatched bars, dashed line) both fit to normal distributions.



**Figure 3 | Kinetic isotope effects.** Values are reported in upright font as mean  $\pm$  s.e.m., with the number of independent trials in parentheses. The second value for the nitrogen nucleophile (asterisked) was previously determined using mass spectrometry instead of scintillation counting as the readout. Calculated values for the transition state in Fig. 4 are shown in italics. Full substrates are shown: cyt, cytidine; cap, caproic acid; bio, biotin; pmn, puromycin.

1.02 to 1.06 have been observed when cleavage of the C–O bond is rate-limiting<sup>22,23</sup>. Conversely, if formation of the amide bond or deprotonation of the T $\pm$  intermediate is the rate-limiting step, then a maximum effect of only 1.01 is expected<sup>22</sup>. We observed a small effect (1.006) on 3'-<sup>18</sup>O substitution, which indicates that there is not significant C–O bond-breaking in the rate-limiting step. This is inconsistent with transition state A, the mechanism in which bond formation to the nucleophile is concerted with breaking the bond to the leaving group, and limits the possible transition states to B, C and D (Fig. 1). Notably, the leaving-group oxygen effect previously predicted for fully concerted proton-shuttle mechanisms varied from 1.023 to 1.038 (ref. 10; transition state A), which is inconsistent with the value of 1.006 measured here.

The primary carbonyl-<sup>13</sup>C isotope effect is derived from a combination of bond breaking (that is, the bond to the leaving group oxygen), bond formation (to the nitrogen) and the accompanying change in hybridization. Given that the bond to the leaving group is intact, the last two factors should dominate. The observed effect for the carbonyl carbon is large and normal, 1.026. For hydrazinolysis of methyl formate, a similar effect has been interpreted as resulting from rate-limiting C–N bond formation<sup>24</sup>. However, it was also shown computationally that an effect of approximately 1.02 could be derived from rate-limiting deprotonation of the zwitterionic intermediate (transition state D)<sup>25</sup>. The remaining isotope effects made it possible to differentiate between these possibilities.

The deuterium effect on the amino acid  $\alpha$ -carbon is primarily derived from the loss of hyperconjugation between the carbonyl  $\pi$ -bond and the antibonding orbital of the C–H bond, which increases the strength of the C–H bond. The magnitude of this effect is largest when the ground state has a large geometric overlap between the  $\pi$ -bond and the antibonding orbital, and the transition state is very close to tetrahedral. Calculations and previous reactions indicate a lower limit of this isotope effect to be approximately 0.96 (refs 21, 26). The measured isotope effect, 0.985, indicates a significant decrease in overlap in the transition state. This is consistent with the decreased  $\pi$ -bond character in a partially tetrahedral transition state.

The KIE of the nucleophilic nitrogen was previously measured to be 1.010, suggestive of an early transition state with partial C–N bond order<sup>18</sup>. We remeasured this value using the remote radiolabel-based assay, and again obtained a normal isotope effect, 1.014. Formation of the C–N bond is expected to produce an inverse isotope effect on the nitrogen. The normal isotope effect observed can be derived from two factors: reaction coordinate motion and deprotonation of the nitrogen. Calculations here and elsewhere indicate that either of these factors alone is insufficient to result in a normal isotope effect (Supplementary Tables 1 and 2, and ref. 10). This suggests that the nitrogen is being deprotonated in the rate-limiting step, simultaneous with C–N bond formation (transition state C). Such a mechanism is consistent with the



near-zero Brønsted coefficient<sup>6</sup>, which indicates that there is no buildup of positive charge on the nucleophile in the transition state.

The 2'-hydroxyl contributes 100- to 2,000-fold to the rate of ribosome-catalysed peptide bond formation<sup>5</sup> and has been postulated to be involved in proton transfer<sup>8–13</sup>. The isotope effect for this position is close to unity (1.002). Changes in bond order between the 2'-oxygen and protons can be compensated by opposite changes in bond strength to the 2'-carbon; therefore this effect is relatively insensitive to the protonation state of the 2'-oxygen<sup>27</sup>. The measured effect indicates that the oxygen does not bear a large positive or negative charge, consistent with previous studies<sup>13</sup>.

To supplement the qualitative descriptions above, we calculated isotope effects for several potential transition state structures (Supplementary Tables 1–3). Although we cannot exclude the possibility that multiple steps contribute to the isotope effects, a single transition state structure did yield calculated values in reasonable agreement with the experimental measurements (Fig. 4). As expected for a transition state, there is a single large imaginary frequency corresponding to reaction coordinate motion (carbon–nitrogen bond formation and nitrogen deprotonation). The error inherent in these calculations and our measurements produces an uncertainty in the precise structure of the transition state; however, a single rate-limiting step is consistent with the measured isotope effects. Formation of the tetrahedral intermediate is simultaneous with deprotonation (transition state C). This step is followed by fast breakdown of the tetrahedral intermediate into products (red pathway in Fig. 1). The data are not consistent with a fully concerted reaction mechanism.

The two-step mechanism identified here is markedly different from what has been observed for similar uncatalysed reactions<sup>7,21</sup>. For uncatalysed reactions at low pH, stepwise formation of T<sup>−</sup> is rate-limiting. At high pH, or with the strong nucleophile hydroxylamine, breakdown of the T<sup>−</sup> intermediate is rate-limiting. Our results indicate that the ribosome not only alters the relative rates of formation and breakdown of tetrahedral intermediates, but also substantially alters the energy landscape such that nucleophilic attack and deprotonation are coordinated.

Concerted attack and deprotonation have also been proposed for aminolysis of the acyl-enzyme intermediate by chymotrypsin<sup>28,29</sup>. This conclusion was supported in part by a low Brønsted coefficient, similar

to that observed for peptide bond formation. This indicates that the  $pK_a$  of the intermediate in the context of these active sites is lower than that of a catalytic group, so that T<sup>±</sup> does not have a finite lifetime and instead T<sup>−</sup> is the only stable intermediate<sup>7</sup>. Although the  $pK_a$  of the amine is initially high, it decreases dramatically as the C–N bond is formed. For chymotrypsin, a catalytic histidine is responsible for deprotonating the nucleophile. For the ribosome, the lack of a pH dependence of the reaction argues against general-acid or general-base catalysis<sup>3</sup>. The 2'-hydroxyl is important for the ribosome reaction, but the  $pK_a$  of a hydroxyl dictates that if it abstracts the amino proton it must donate its proton elsewhere. This mechanism, termed the 'proton shuttle', has been proposed in a variety of forms for the ribosome<sup>8–12</sup>. Alternatively, the 2'-hydroxyl may provide an important transition state hydrogen bond, with another group or a water molecule deprotonating the nucleophile. Either possibility would be consistent with measured solvent isotope effects, which indicate more than one proton is in motion at the transition state (S. Kuhlenkoetter and M. Rodnina, personal communication).

The destination of the proton originating from the nucleophile is uncertain. Two computational studies favour the carbonyl oxygen, which would otherwise develop a negative charge<sup>11,12</sup>. This would result in an uncharged tetrahedral intermediate, which would be consistent with the lack of pH dependence of the reaction. Alternatively, a water molecule observed in crystal structures is correctly positioned to accept a proton from the 2'-hydroxyl and later donate it to the 3'-oxygen leaving group. Finally, the 3'-oxygen could receive the proton from the 2'-hydroxyl. Our data cannot distinguish between these possibilities; for simplicity, we have modelled a water molecule as the proton acceptor.

The ribosome increases the rate of peptide bond formation by an estimated 10<sup>7</sup>-fold through an altered chemical mechanism. The rate of nucleophilic attack is probably increased through positioning effects—first, the increased likelihood of the two substrates being in proximity, and second, precise orientation of the nucleophile (possibly by the 2'-hydroxyl of A2451; ref. 2). The ribosome also has a significant catalytic role, beyond substrate positioning, by coordinating nucleophilic attack and deprotonation in a single rate-limiting step. Catalysis by the ancient, conserved, RNA active site of the ribosome fundamentally alters the reaction pathway for peptide bond formation relative to the uncatalysed reaction.

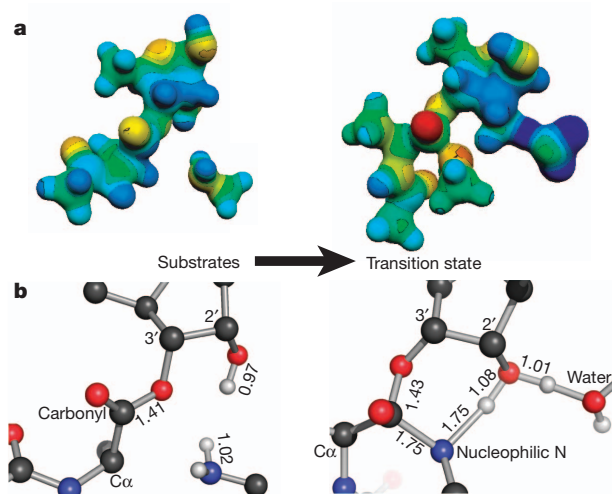
## METHODS SUMMARY

A general strategy for measuring isotope effects with these substrates using a competitive assay has been described<sup>21</sup>. The P-site substrate cytidyl-(3'5')-cytidyl-(3'5')-3'(2')-O-(N-(6-D-(+)-biotinoylamino)hexanoyl)-L-phenylalanyl adenosine (CCApbc) was synthesized as described<sup>15,16</sup>. For each measurement, two substrates differing by a single isotopic substitution at the reaction centre were mixed. Each substrate was labelled with either <sup>32</sup>P or <sup>33</sup>P to allow differentiation by scintillation counting.

Reaction mixtures contained 7 mM MgCl<sub>2</sub>, 140 mM NH<sub>4</sub>Cl, 25 mM HEPES pH 8.5, trace P-site substrate, 250–500 mM A-site substrate, and 6–10 μM 50S ribosomes. At these concentrations, the isotope effect is on  $k_{cat}/K_M$  for the P-site substrate. Aliquots were quenched at 15–30% reacted and near the reaction end-point. Substrate and product were separated on an acrylamide gel and the fraction reacted determined by phosphorimaging. Each band was eluted and scintillation counted to determine the ratio of <sup>32</sup>P to <sup>33</sup>P in each sample.

To account for substrate that hydrolysed before the reaction was started, the amount of <sup>32</sup>P- and <sup>33</sup>P-labelled product at zero time was subtracted from all determinations of the product isotope ratio. Additionally, some substrate was unreactive even at long time points. The amounts of <sup>32</sup>P- and <sup>33</sup>P-labelled unreactive substrate were similarly subtracted from all determinations of the substrate isotope ratio.

Isotope effects were calculated as previously described<sup>21</sup>, using hybrid density functional methods implemented in Gaussian03. Transition state structures were optimized and frequencies computed on the optimized structures using the three-parameter Becke exchange functional, the LYP correlation functional and the standard 6-31G(d,p) basis set. Isotope effects were calculated from the computed frequencies using ISOEFF 98<sup>30</sup>. Methylamine was used to model the nucleophile. Calculated isotope effects for tetrahedral intermediate formation were repeated



**Figure 4 | Structures of the substrates and transition states for peptide bond formation shown in similar orientations.** **a**, Electrostatic representations. Blue is electron-deficient (cationic) and red is electron-rich (anionic). **b**, Geometric representations. Carbons are shown grey, oxygens are red, nitrogens are blue, and hydrogens are white. Protons on the nucleophilic nitrogen, 2'-hydroxyl and water are displayed; for clarity, all others are omitted. Bond lengths are shown in Å. The structure was obtained by matching isotope effects calculated with Gaussian03 and Isoeff to the experimental values.

with a set of nucleophiles of increasing size and complexity; these effects varied by less than 0.003.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 28 April; accepted 3 June 2011.**

**Published online 17 July 2011.**

- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Erlacher, M. D. *et al.* Chemical engineering of the peptidyl transferase center reveals an important role of the 2'-hydroxyl group of A2451. *Nucleic Acids Res.* **33**, 1618–1627 (2005).
- Bieling, P., Beringer, M., Adio, S. & Rodnina, M. V. Peptide bond formation does not involve acid-base catalysis by ribosomal residues. *Nature Struct. Mol. Biol.* **13**, 423–428 (2006).
- Sievers, A., Beringer, M., Rodnina, M. V. & Wolfenden, R. The ribosome as an entropy trap. *Proc. Natl Acad. Sci. USA* **101**, 7897–7901 (2004).
- Zaher, H. S., Shaw, J. J., Strobel, S. A. & Green, R. The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *EMBO J.* **30**, 2445–2453 (2011).
- Kingery, D. A. *et al.* An uncharged amine in the transition state of the ribosomal peptidyl transfer reaction. *Chem. Biol.* **15**, 493–500 (2008).
- Satterthwait, A. C. & Jencks, W. P. Mechanism of the aminolysis of acetate esters. *J. Am. Chem. Soc.* **96**, 7018–7031 (1974).
- Dorner, S., Panuschka, C., Schmid, W. & Barta, A. Mononucleotide derivatives as ribosomal P-site substrates reveal an important contribution of the 2'-OH to activity. *Nucleic Acids Res.* **31**, 6536–6542 (2003).
- Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A. & Steitz, T. A. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol. Cell* **20**, 437–448 (2005).
- Wallin, G. & Åqvist, J. The transition state for peptide bond formation reveals the ribosome as a water trap. *Proc. Natl Acad. Sci. USA* **107**, 1888–1893 (2010).
- Wang, Q., Gao, J., Liu, Y. & Liu, C. Validating a new proton shuttle reaction pathway for formation of the peptide bond in ribosomes: a theoretical investigation. *Chem. Phys. Lett.* **501**, 113–117 (2010).
- Rangelov, M. A., Petrova, G. P., Yomtova, V. M. & Vayssilov, G. N. Catalytic role of vicinal OH in ester aminolysis: proton shuttle versus hydrogen bond stabilization. *J. Org. Chem.* **75**, 6782–6792 (2010).
- Huang, K. S., Carrasco, N., Pfund, E. & Strobel, S. A. Transition state chirality and role of the vicinal hydroxyl in the ribosomal peptidyl transferase reaction. *Biochemistry* **47**, 8822–8827 (2008).
- Cleland, W. W. Isotope effects: determination of enzyme transition state structure. *Enzyme Kinetics Mech. D* **249**, 341–373 (1995).
- Zhong, M. & Strobel, S. A. Synthesis of the ribosomal P-site substrate CCA-pcb. *Org. Lett.* **8**, 55–58 (2006).
- Zhong, M. & Strobel, S. A. Synthesis of isotopically labeled P-site substrates for the ribosomal peptidyl transferase reaction. *J. Org. Chem.* **73**, 603–611 (2008).
- Monro, R. E. & Marcker, K. A. Ribosome-catalysed reaction of puromycin with a formylmethionine-containing oligonucleotide. *J. Mol. Biol.* **25**, 347–350 (1967).
- Seila, A. C., Okuda, K., Nunez, S., Seila, A. F. & Strobel, S. A. Kinetic isotope effect analysis of the ribosomal peptidyl transferase reaction. *Biochemistry* **44**, 4018–4027 (2005).
- Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
- Unrau, P. J. & Bartel, D. P. An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. *Proc. Natl Acad. Sci. USA* **100**, 15393–15397 (2003).
- Hiller, D. A., Zhong, M., Singh, V. & Strobel, S. A. Transition states of uncatalyzed hydrolysis and aminolysis reactions of a ribosomal P-site substrate determined by kinetic isotope effects. *Biochemistry* **49**, 3868–3878 (2010).
- O'Leary, M. H. & Marlier, J. F. Heavy-atom isotope effects on the alkaline hydrolysis and hydrazinolysis of methyl benzoate. *J. Am. Chem. Soc.* **101**, 3300–3306 (1979).
- Sawyer, C. B. & Kirsch, J. F. Kinetic isotope effects for reactions of methyl formate-methoxyl-<sup>18</sup>O. *J. Am. Chem. Soc.* **95**, 7375–7381 (1973).
- Marlier, J. F., Haptonstall, B. A., Johnson, A. J. & Sacksteder, K. A. Heavy-atom isotope effects on the hydrazinolysis of methyl formate. *J. Am. Chem. Soc.* **119**, 8838–8842 (1997).
- Singleton, D. A. & Merrigan, S. R. Resolution of conflicting mechanistic observations in ester aminolysis. A warning on the qualitative prediction of isotope effects for reactive intermediates. *J. Am. Chem. Soc.* **122**, 11035–11036 (2000).
- Hogg, J. L., Rodgers, J., Kovach, I. & Schowen, R. L. Kinetic isotope-effect probes of transition-state structure. Vibrational analysis of model transition states for carbonyl addition. *J. Am. Chem. Soc.* **102**, 79–85 (1980).
- Gawlita, E. *et al.* H-Bonding in alcohols is reflected in the C $\alpha$ –H bond strength: variation of C–D vibrational frequency and fractionation factor. *J. Am. Chem. Soc.* **122**, 11660–11669 (2000).
- Inward, P. W. & Jencks, W. P. The reactivity of nucleophilic reagents with furoyl-chymotrypsin. *J. Biol. Chem.* **240**, 1986–1996 (1965).
- Zeeberg, B. & Caplow, M. Transition state charge distribution in reactions of an acetyltyrosylchymotrypsin intermediate. *J. Biol. Chem.* **248**, 5887–5891 (1973).
- Anisimov, V. & Paneth, P. ISOEFF98. A program for studies of isotope effects using Hessian modifications. *J. Math. Chem.* **26**, 75–86 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Klinman and D. Singleton for critical reading of the manuscript, and members of the S.A.S. laboratory for discussion. This work was supported by an NIH postdoctoral fellowship (D.A.H.), a Brown-Coxe fellowship (V.S.) and an NIH grant (GM54839).

**Author Contributions** D.A.H. and S.A.S. devised experiments, M.Z. synthesized substrates, D.A.H. collected and analysed isotope effect data, V.S. calculated theoretical isotope effects, and D.A.H. and S.A.S. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.A.S. ([scott.strobel@yale.edu](mailto:scott.strobel@yale.edu)).

## METHODS

**Reaction assay.** A general strategy for measuring isotope effects with these substrates using a competitive assay has been described<sup>21</sup>. The P-site substrate cytidyl-(3'5')-cytidyl-(3'5')-3'(2')-O-(N-(6-D-(+)-biotinoylamino)hexanoyl)-L-phenylalanyl)adenosine (CCApb) was synthesized as previously described<sup>15,16</sup>. Substrates were prepared that differed only by a single isotopic substitution at the reaction centre. Each substrate was 5'-end labelled with either  $\gamma^{32}\text{P}$ - or  $\gamma^{33}\text{P}$ -ATP by polynucleotide kinase and purified on a pH 6 denaturing 15% polyacrylamide gel.  $^{32}\text{P}$ - and  $^{33}\text{P}$ -CCApb were mixed at an approximately 1:4 ratio and purified on a pH 6 non-denaturing 12% polyacrylamide gel. Purified mixes were used as soon as possible to minimize hydrolysis in the starting material.

The isotopic enrichment of each sample was determined by high-resolution Fourier-transform-ion cyclotron resonance mass spectrometry. Each molecule had 95% or greater isotopic purity.

Ribosome reaction mixtures contained 7 mM  $\text{MgCl}_2$ , 140 mM  $\text{NH}_4\text{Cl}$ , 25 mM HEPES pH 8.5, trace P-site substrate, 250–500 mM A-site substrate, and 6–10  $\mu\text{M}$  50S ribosomes. At these concentrations the A-site substrate is saturating but the ribosome concentration is not; therefore the isotope effect is on  $k_{\text{cat}}/K_{\text{M}}$  for the P-site substrate. All reactions were performed at room temperature. Reaction aliquots at 15–30% reacted (approximately 1 min) and near the reaction endpoint (greater than 30 min) were quenched with three volumes of formamide loading buffer. Each time point was run on a pH 6 denaturing 15% polyacrylamide gel to separate substrate from product.

Substrate and product bands were visualized using a Storm 840 PhosphorImager with a two-ply sheet of Duck tape between the gel and screen to block  $^{33}\text{P}$  emission. The fraction of  $^{32}\text{P}$ -labelled substrate reacted could then be determined. The ratio of  $^{32}\text{P}$  to  $^{33}\text{P}$  was determined by scintillation counting. Each product band was excised from the gel and eluted into 1 ml of 50 mM NaCl overnight. This elution was then added to 13 ml of Optima Gold scintillation fluid and counted for 30 min along with  $^{32}\text{P}$  and  $^{33}\text{P}$  standards.

**Data analysis.** Counts per minute were divided into two channels, 0–400 keV and 400–2,000 keV. Approximately 70% of the  $^{32}\text{P}$  standard was detected in the high energy channel, and greater than 99% of the  $^{33}\text{P}$  sample was detected in the low energy channel. The ratio of  $^{32}\text{P}$  to  $^{33}\text{P}$  in each sample could be determined using equation (1):

$$^{33}\text{P}/^{32}\text{P} = (A - Br)/(B(1 + r)) \quad (1)$$

where  $A$  is the counts per minute in the low energy channel,  $B$  is the counts per minute in the high energy channel, and  $r$  is the ratio of emission of a  $^{32}\text{P}$  standard detected in the low energy channel to the high energy channel. To account for substrate that hydrolysed before the reaction was started, the amount of  $^{32}\text{P}$ - and  $^{33}\text{P}$ -labelled product at zero time was determined. These were subtracted from all other determinations of the product isotope ratio. Additionally, some substrate was unreactive even at long time points. The amounts of  $^{32}\text{P}$ - and  $^{33}\text{P}$ -labelled unreactive substrate were similarly subtracted from all other determinations of the substrate isotope ratio. The observed isotope effect was then determined from the ratio in the midpoint and endpoint samples and the fraction reacted using equation (2):

$$\text{KIE} = \log(1 - f)/\log(1 - fR_p/R_0) \quad (2)$$

where  $f$  is the fraction reacted,  $R_p$  is the isotope ratio in the product at that fraction reacted, and  $R_0$  is the isotope ratio in the product at the reaction endpoint. For

high-fraction-reacted samples (greater than 50%), the isotope effect was also determined from substrate and endpoint samples using equation (3):

$$\text{KIE} = \log(1 - f)/\log[(1 - f)R_s/R_0] \quad (3)$$

where  $R_s$  is the isotope ratio in the remaining substrate at  $f$  fraction reacted, and  $R_0$  is as above. This value was corrected for incomplete isotopic incorporation as determined by mass spectrometry with equation (4):

$$\text{KIE}_{\text{corrected}} = 1 + (\text{KIE}_{\text{observed}} - 1)/[1 - \text{KIE}_{\text{observed}}(1 - e)] \quad (4)$$

where  $e$  is the isotopic enrichment of the heavy sample. This equation assumes a negligible amount of heavy isotope in the light sample<sup>14,21</sup>.

For the control and  $\alpha$ -deuterium isotope effects, enough trials were performed to plot a histogram of the data (Fig. 2). Both data sets were fitted to a normal distribution, equation (5):

$$c = a \exp\{-0.5[(x - \mu)/\sigma]^2\} \quad (5)$$

where  $c$  is the number of counts in a given bin,  $a$  is the amplitude,  $\mu$  is the mean and  $\sigma$  is the standard deviation. For some substitutions, the standard errors for trials performed on the same day were smaller than those from different days, indicating that each trial may not be independent. Therefore, for all substitutions each set of experiments from a single day were averaged. The mean and standard error of multiple data sets are reported.

**Computation of transition states.** Transition state structures that reproduced the experimental KIEs were determined using hybrid density functional methods implemented in Gaussian03. Tetrahydro-4,5-dihydroxy-2-(methoxymethyl)furan-3-yl 2-acetamidopropanoate was used to model the P-site substrate and methylamine was used to model the A-site nucleophile. Calculated isotope effects for tetrahedral intermediate formation were repeated with a set of nucleophiles of increasing size and complexity; the calculated isotope effects varied by less than 0.003.

Structures of the transition states were optimized and the frequencies were computed for the optimized structures using the three-parameter Becke (B3) exchange functional, the LYP correlation functional and the standard 6-31G(d,p) basis set. The 5'-methoxy group and the reaction centre were constrained during the optimization and many of these constraints were modified to match the experimental KIEs.

KIEs and equilibrium isotope effects (EIEs) were calculated from the computed frequencies using ISOEFF 98<sup>30</sup>. KIEs and EIEs were calculated for a temperature of 298 K and the frequencies were scaled using a factor of 0.964, corresponding to B3LYP/6-31G(d,p). KIEs were calculated whenever the magnitude of the imaginary frequency was greater than  $50 \text{ i cm}^{-1}$ , otherwise EIEs were computed. All vibrational modes were used to calculate isotope effects.

Geometric and the electrostatic models were generated by iteratively optimizing the transition states by modifying the applied constraints until the computed isotope effects closely matched the experimental KIEs.

The natural bond orbital (NBO) calculations were performed on optimized structures by including the `pop = nbo` keyword in the route section of input files. The molecular electrostatic potential (MEP) surfaces were calculated by the CUBE subprogram of Gaussian03. The formatted checkpoint files used in the CUBE subprogram were generated by constrained geometry optimization at the B3LYP level of theory with the 6-31G\*\* basis set. MEP surfaces of the substrate and the transition states were visualized using Molekel4.0 at a density of 0.2 electrons per  $\text{\AA}^3$ . Geometric figures were created using PyMol.



# Collaboration encourages equal sharing in children but not in chimpanzees

Katharina Hamann<sup>1</sup>, Felix Warneken<sup>2</sup>, Julia R. Greenberg<sup>3</sup> & Michael Tomasello<sup>1</sup>

**Humans actively share resources with one another to a much greater degree than do other great apes, and much human sharing is governed by social norms of fairness and equity<sup>1–3</sup>. When in receipt of a windfall of resources, human children begin showing tendencies towards equitable distribution with others at five to seven years of age<sup>4–7</sup>. Arguably, however, the primordial situation for human sharing of resources is that which follows cooperative activities such as collaborative foraging, when several individuals must share the spoils of their joint efforts<sup>8–10</sup>. Here we show that children of around three years of age share with others much more equitably in collaborative activities than they do in either windfall or parallel-work situations. By contrast, one of humans' two nearest primate relatives, chimpanzees (*Pan troglodytes*), 'share' (make food available to another individual) just as often whether they have collaborated with them or not. This species difference raises the possibility that humans' tendency to distribute resources equitably may have its evolutionary roots in the sharing of spoils after collaborative efforts.**

Among great apes, only humans are true collaborative foragers<sup>8,9,11</sup>. Other apes forage in small parties, but they do not actively work together jointly to produce food—the only exception being chimpanzee group-hunting of monkeys<sup>12,13</sup>. In contrast, humans in all societies produce significant portions of their food through collaborative efforts, even bringing the results of their labour back to some central location to share with other group members<sup>14,15</sup>. After group-hunting, chimpanzees mostly share only under pressure of harassment by others<sup>16</sup> or else reciprocally with coalition partners<sup>17</sup>.

Human children actively share valuable resources with others to some degree from early in ontogeny. A fairly well-established pattern across cultures is that three- to four-year-old children tend to divide a windfall of resources unequally, keeping the majority for themselves<sup>4–6,18,19</sup>. As they approach school age, they begin to share more equally<sup>4,5,7,18,19</sup>. But given that humans generate many or most of their resources collaboratively, a plausible hypothesis is that children would share a resource more equitably at an earlier age if it was not provided by adults as a windfall, but if instead they had to work together to produce it<sup>20</sup>. Furthermore, we might expect this positive effect of collaboration on sharing to be confined to humans, among great apes, as only they have an evolutionary history of obligate collaborative foraging<sup>8,9,11</sup>.

In the current series of experiments, therefore, we presented pairs of human children and pairs of chimpanzees with resource distribution problems in which one individual had control of more than half of the resources and could choose whether or not to share them equally with their partner. The basic variable was whether the initial unequal distribution of resources resulted from a collaborative effort in which each contributed equally, or whether it came from some non-collaborative source (for example as a windfall or as a result of each individual working on their own).

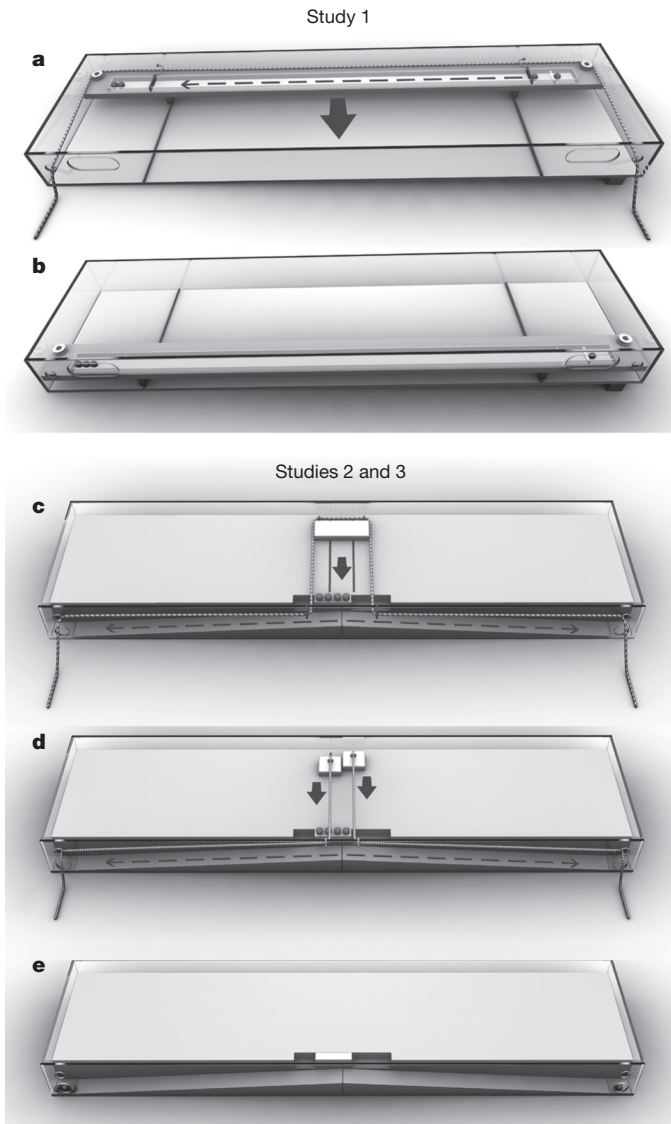
In study 1, pairs of either two- or three-year-old children were in a room by themselves. In the 'collaboration' condition, they faced an enclosed board with a rope extruding from each end (Fig. 1a), and they knew from previous experience (from a demonstration phase) that

they had to pull together to bring the board towards them. On each end of the board were two rewards (small toys) that could be accessed once the board had been pulled close enough. As the children pulled, one of the toys rolled to the other end of the board such that one child ended up with three toys and the other ended up with only one. In the control, 'no-work', condition, by contrast, as children entered the room the board with the toys was already at its end-state position, with three toys at one end and one at the other (Fig. 1b). The main result was that the 'lucky' child, who had gained three toys, made one of the toys available to the 'unlucky' partner, who had gained one, restoring equity, more often in the collaboration condition than in the no-work condition ( $F(1, 22) = 21.85$  (analysis of variance),  $P < 0.001$ ). The effect was similar for children of both ages (Fig. 2a).

In this experiment, it was possible that from the beginning of the collaboration children viewed the rewards on their end of the board as belonging to them, such that when one reward rolled to the other end it was as if one of their possessions had been taken away (which was not the case in the no-work condition). In study 2, therefore, we presented pairs of two- or three-year-old children initially with four toys bunched together, so that an initial sense of possession was not an issue. In addition, we added a second control condition—the parallel-work condition—with a very similar set-up, in which each child pulled on a separate board with their own separate rope, to account for the fact that the collaboration condition required work whereas the original control condition (no-work) did not (Fig. 1c–e). Thus, if children are attentive to work effort in general and not to collaborative effort in particular, they should share similarly in the parallel-work and collaboration conditions. However, in this study also, the three-year-old lucky child handed over one of the toys to the unlucky partner more often in the collaboration condition than in either of the two control conditions (no-work and parallel-work). By contrast, the two-year-olds did not differentiate among conditions (see Fig. 2b for data and statistics for both ages).

Because studies 1 and 2 consisted of multiple trials, they leave open the possibility that children shared in the collaboration condition out of a concern that if they did not share their partner might not pull their end of the rope in future trials (which was not an issue in the parallel-work and no-work conditions as children obtained rewards on their own.) In study 3, to ensure that children understood that they would play the game only once, in the demonstration phase we showed them the total number of toys available and made it clear that their number decreased over demonstrations. When there was only one set of four toys left, we pointed this out and specifically asked the children whether the game could be played after this last set was gone. Only children who answered that this would be the last time were then given the actual test trial (that is, only their data was used for analysis; see Supplementary Information for details). Replicating our results once more, three-year-old children equalized the distribution of toys more often in the collaboration trials (75%) than in the parallel-work trials (25%;  $\chi^2_{(d.f.=1, n=24)} = 6.0$ ,  $P = 0.039$ ; Fig. 2c). Taken together, these studies show that collaborative work encourages equal sharing in

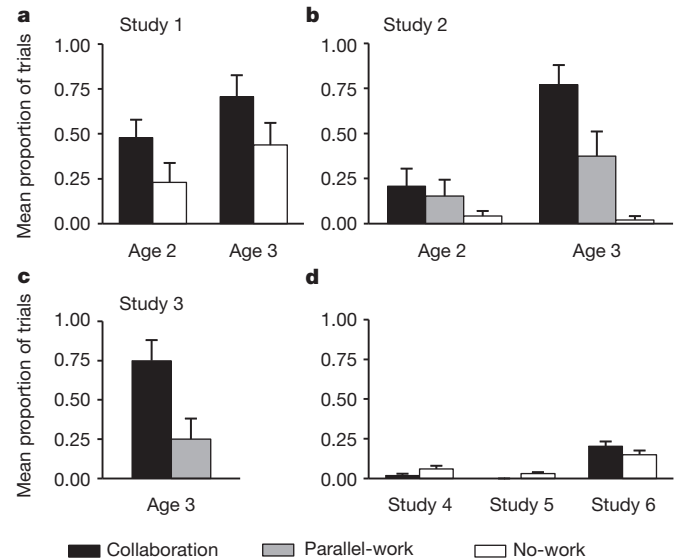
<sup>1</sup>Max Planck Institute for Evolutionary Anthropology, Department of Developmental and Comparative Psychology, Deutscher Platz 6, 04103 Leipzig, Germany. <sup>2</sup>Harvard University, Department of Psychology, 33 Kirkland St, Cambridge, Massachusetts 02138, USA. <sup>3</sup>Michigan State University, Department of Zoology, East Lansing, Michigan 48824-1115, USA.



**Figure 1 | Child study tasks.** **a**, Apparatus from study 1 with reward relocation mechanism as used in the collaboration condition (180 cm × 60 cm × 15 cm; adapted from studies with chimpanzees<sup>29,30</sup>). In the collaboration condition, children had to pull both ends of the rope simultaneously to move the board towards the access holes in the front of the enclosure (solid arrow). Initially, two toys (marbles) were on each side (as shown), but as the children pulled the board closer, the black barriers slipped out such that one marble rolled to the other end, resulting in a 3:1 reward distribution (moving from right to left in this example; dashed arrow). **b**, In the no-work condition, the board was already in the front part of the apparatus, with no attached rope, when children approached it (same reward distribution, of 3:1). **c**, In studies 2 and 3, children had to move a block closer to move the marbles such that they would roll in front of the access holes. In the collaboration condition, children had to pull a single, long rope simultaneously to move a large block closer (solid arrow), moving four marbles at once, which then rolled towards the respective access holes (in this example, three marbles rolled to the left and one marble rolled to the right; dashed arrows). **d**, In the parallel-work condition, two smaller blocks (each with a rope attached) could be pulled individually, one by each child, causing the respective marbles to move and roll down the ramps. **e**, The no-work condition, without any work but with the same reward distribution, 3:1.

children much more than does working in parallel or acquiring resources in a windfall.

Chimpanzees do not regularly offer resources to others actively, so to test for the same effect of collaboration on sharing in chimpanzees we had to use a slightly more complex apparatus that enabled one individual to provide another with food that the second could not

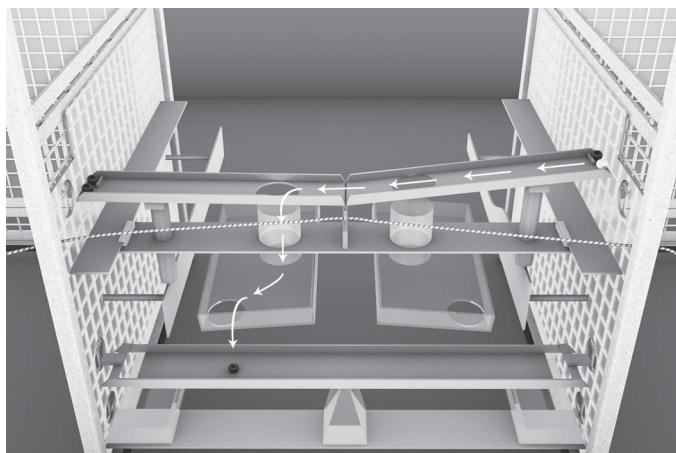


**Figure 2 | Rates of equal shares.** **a**, In study 1, children in both age groups shared more often in the collaboration condition than in the no-work condition ( $F(1, 22) = 21.85, P < 0.001$ ). This was true even if only the results of the first trial in both conditions were used for the analysis (McNemar test,  $n_{1,0} = 0$  (number of dyads sharing in the no-work condition but not in the collaboration condition),  $n_{0,1} = 6$  (number of dyads sharing in the collaboration condition but not in the no-work condition),  $P = 0.03$ ). **b**, In study 2, three-year-olds, but not two-year-olds, shared differently in the three conditions (significant age × condition interaction,  $F(2, 66) = 5.26, P = 0.008$ ; main effect of condition,  $F(2, 66) = 12.87, P < 0.001$ ). Three-year-olds shared significantly more often in the collaboration condition than in either of the other two conditions (*post hoc* Scheffé tests, both  $P < 0.05$ ). The difference between the parallel-work and no-work conditions approached significance ( $P = 0.06$ ). **c**, In study 3, children shared significantly more often in the collaboration condition as compared with the parallel-work condition ( $\chi^2_{(d.f.=1, n=24)} = 6.0, P = 0.039$ ). **d**, Across studies 4–6, chimpanzees did not share differently in the collaboration and control (no-work) conditions. See main text and Supplementary Information for details and additional analyses. Error bars, s.e.m.

obtain. Although some researchers have proposed that chimpanzees use work effort during group hunts as a criterion for dividing up the spoils<sup>21</sup>, our hypothesis was that because chimpanzees are not true collaborative foragers (at least not to the degree of humans<sup>8,9,13</sup>), they would not share differently in windfall and collaboration situations.

The two chimpanzees operated a single apparatus but from adjacent rooms (after enough practice with the apparatus for both to know how it worked; Supplementary Information). The upper level of the apparatus (Fig. 3) was similar to the apparatus used in the experiment with children, in that it contained a long board holding rewards (in this case food) and attached to a rope that each chimpanzee could access. In all conditions, at some point there was one piece of food on a lower level on a see-saw device, such that the lucky chimpanzee could tip the see-saw only either towards itself or away from itself and towards its partner. In a series of three experiments, we made it increasingly easy for one chimpanzee to make the fallen piece of food available to its partner. In all three experiments, there was a collaboration condition, in which the pair worked together to pull on the board, and a control (windfall) condition, in which the food got into position without the chimpanzees' joint effort.

The first chimpanzee study (study 4) was very similar to the first child study (as this procedure was most facilitative of sharing for the children). In this study, the lucky chimpanzee (who got two pieces of food and had a chance to get the fallen piece) could take the fallen piece for itself or could restore the 2:2 balance by actively providing the fallen piece to the unlucky partner (who got one piece of food and had a chance to get the fallen piece) or by doing nothing and letting the unlucky partner tip it to itself. What happened most often was that



**Figure 3 | Chimpanzee study task.** The apparatus was mounted in a booth between testing rooms. Two chimpanzees located in adjacent rooms (partition mesh not shown) had to pull the two ends of the rope simultaneously to move the upper-level platform in front of the access holes. During the movement, one of the unlucky individual's rewards (grapes) fell onto the lower, see-saw, level and rolled to the lucky chimpanzee's side. In the picture, the sliding platform (upper level) has been pulled almost to the front. This movement has caused one side of the platform to tilt, causing one grape from the right (initial location represented by the grape drawn in white) to roll to the left, falling through a hole and landing on the see-saw mechanism below. The see-saw could then be tilted by the chimpanzees (manner dependent on study and condition).

the unlucky partner almost immediately tipped the see-saw and took the fallen reward for itself (63% of trials); in no cases did the lucky chimpanzee actively tip the reward to the unlucky partner (even though in pre-training they often tipped the food away from themselves into the other room, if they themselves could then go through an open door and get it). In the remaining trials, the lucky chimpanzee took the reward for itself. Importantly, the fate of the fallen reward did not differ between conditions (Wilcoxon signed-rank test,  $T+ = 31.5$ ,  $n = 12$  (no ties),  $P = 0.60$ ). Additional analyses showed that in only 4% of cases did the lucky chimpanzee in fact give up the food voluntarily by tolerating the unlucky partner's taking of it (again with no difference between conditions:  $T+ = 18$ ,  $n = 11$  (one tie),  $P = 0.19$ , Fig. 2d).

In the second chimpanzee study (study 5), we tried to encourage the lucky partner by making it impossible for the unlucky partner to operate the see-saw. The result was that the lucky chimpanzee almost always tipped the food to itself (98% of trials), thus creating a 3:1 reward imbalance. There was again no difference in the rate of equitable sharing between the collaboration and control conditions (Wilcoxon signed-rank test,  $T+ = 0$ ,  $n = 4$  (eight ties),  $P = 0.13$ ). In the third chimpanzee study (study 6), we made it impossible for the lucky partner to get the fallen piece of food for itself: if it tipped the see-saw towards itself, the food was lost. This new set-up resulted in a higher sharing rate than before (mean, 0.17 of trials; s.d. = 0.17), with subjects tipping the reward to the unlucky partner more often than in both chimpanzee study 1 ( $T+ = 69$ ,  $n = 12$  (no ties),  $P = 0.016$ ) and chimpanzee study 2 ( $T+ = 64$ ,  $n = 11$  (one tie),  $P = 0.003$ ). However, as in the previous two chimpanzee experiments, the results did not differ between the collaboration and control conditions ( $T+ = 40$ ,  $n = 10$  (two ties),  $P = 0.22$ ; Fig. 2d).

Previous research with older school-age children (seven to ten years of age) has shown that they take into account work effort in so-called distributive justice problems in which one individual must say how the fruits of a collaborative effort should be doled out to participants<sup>22–26</sup>. Younger children typically are not able to factor work effort into their decisions in this way. Nevertheless, the current study shows that although they may be unable to balance work and rewards sensitively, children as young as two or three years of age do take note of whether

or not rewards were produced from collaborative efforts with others, and that this affects how they think the rewards should be distributed. Thus, the ontogenetically first sense of distributive justice may be that participation in a collaborative effort demands an equal division of spoils. Because chimpanzees rely very little on collaboration for subsistence, they have not evolved the tendency to distribute resources more equally when those resources result from a collaboration.

Collaborative foraging, by definition, requires partners. Any individual with a tendency to take more than their share of the fruits of a collaboration would not be chosen as a partner very often<sup>27,28</sup>. A possible evolutionary picture is thus that this 'social selection' of a tendency to share the fruits of collaboration equally among participants became ever stronger as the need to work together jointly in subsistence activities became ever more obligate. The current results, according to which young children, but not chimpanzees, share more equally after collaboration than in other situations, provide at least indirect support for this picture.

## METHODS SUMMARY

**Children.** We tested children at 2 and 3 years of age (study 1,  $n = 48$ ; study 2,  $n = 144$ ; study 3,  $n = 48$ ) who were paired with a same-sex peer from the same kindergarten. In the demonstration phase of each study, we first familiarized children with the apparatus requiring them to pull ropes to retrieve rewards (marbles to play an individual game). In all studies, the test event was that the lucky child ended up with three marbles and the unlucky child ended up with only one.

In study 1, during the demonstration phase both individuals learned how to pull together to bring an enclosed board holding rewards within reach of access holes in the enclosure (Fig. 1a). In the test phase, we presented two conditions, collaboration and no-work, in counterbalanced order within dyads.

In study 2, different pairs of children worked on a slightly modified version of the original apparatus (Fig. 1c). Children were tested either in a collaboration condition (preceded by a collaborative demonstration as in study 1), a parallel-work condition (preceded by an individual work demonstration; Fig. 1d) or a no-work condition (preceded by a joint no-work demonstration; Fig. 1e).

In study 3, pairs of three-year-old children participated either in a single collaboration test trial or in a single parallel-work test trial (preceded by demonstrations similar to those in study 2.)

**Chimpanzees.** We tested 12 chimpanzees separately with two partners from their social group, for a total of 12 test pairs. We conducted three studies with increasing levels of encouragement of the lucky individual to share. We achieved this by blocking the holes the chimpanzees could use to tip the see-saw or retrieve the food (Fig. 3). In all three experiments, we presented the collaboration and no-work conditions in a within-subject design administered in counterbalanced order (for details, see Supplementary Information).

Received 12 September 2010; accepted 9 June 2011.

Published online 20 July 2011.

1. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
2. Gurven, M. To give or not to give: an evolutionary ecology of human food transfers. *Behav. Brain Sci.* **27**, 543–583 (2004).
3. Henrich, J. et al. "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–855 (2005).
4. Benenson, J. F., Pascoe, J. & Radmore, N. Children's altruistic behavior in the dictator game. *Evol. Hum. Behav.* **28**, 168–175 (2007).
5. Fehr, E., Bernhard, H. & Rockenbach, B. Egalitarianism in young children. *Nature* **454**, 1079–1083 (2008).
6. Blake, P. R. & Rand, D. G. Currency value moderates equity preference among young children. *Evol. Hum. Behav.* **31**, 210–218 (2010).
7. Gummerum, M., Hanoch, Y., Keller, M., Parsons, K. & Hummel, A. Preschoolers' allocations in the dictator game. The role of moral emotions. *J. Econ. Psychol.* **31**, 25–34 (2010).
8. Tomasello, M. *Why We Cooperate* (MIT Press, 2009).
9. Sterelny, K. Social intelligence, human intelligence and niche construction. *Phil. Trans. R. Soc. B* **362**, 719–730 (2007).
10. de Waal, F. B. M. & Berger, M. L. Payment for labour in monkeys. *Nature* **404**, 563 (2000).
11. Hill, K. Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Hum. Nature* **13**, 105–128 (2002).
12. Boesch, C. & Boesch, H. Hunting behavior of wild chimpanzees in the Tai-National Park. *Am. J. Phys. Anthropol.* **78**, 547–573 (1989).
13. Muller, M. & Mitani, J. C. in *Advances in the Study of Behavior* (eds Slater, P. J. B., Rosenblatt, J., Snowdon, C., Roper, T. & Naguib, M.) 275–331 (Elsevier, 2005).



14. Hill, K., Barton, M. & Hurtado, A. M. The emergence of human uniqueness: characters underlying behavioral modernity. *Evol. Anthropol.* **18**, 187–200 (2009).
15. Marlowe, F. W. Hunter-gatherers and human evolution. *Evol. Anthropol.* **14**, 54–67 (2005).
16. Gilby, I. C. Meat sharing among the Gombe chimpanzees: harassment and reciprocal exchange. *Anim. Behav.* **71**, 953–963 (2006).
17. Mitani, J. C. & Watts, D. Why do chimpanzees hunt and share meat? *Anim. Behav.* **61**, 915–924 (2001).
18. Thompson, C., Barresi, J. & Moore, C. The development of future-oriented prudence and altruism in preschoolers. *Cogn. Dev.* **12**, 199–212 (1997).
19. Rochat, P. *et al.* Fairness in distributive justice by 3- and 5-year-olds across seven cultures. *J. Cross Cult. Psychol.* **40**, 416–442 (2009).
20. Warneken, F., Lohse, K., Melis, A. P. & Tomasello, M. Young children share the spoils after collaboration. *Psychol. Sci.* **22**, 267–273 (2011).
21. Boesch, C. Cooperative hunting in wild chimpanzees. *Anim. Behav.* **48**, 653–667 (1994).
22. Lerner, M. J. The justice motive: “equity” and “parity” among children. *J. Pers. Soc. Psychol.* **29**, 539–550 (1974).
23. Damon, W. *The Social World of the Child* (Jossey-Bass, 1977).
24. Hook, J. G. & Cook, T. D. Equity theory and the cognitive ability of children. *Psychol. Bull.* **86**, 429–445 (1979).
25. Kienbaum, J. & Wilkening, F. Children’s and adolescents’ intuitive judgments about distributive justice: integrating need, effort, and luck. *Eur. J. Dev. Psychol.* **6**, 481–498 (2009).
26. Almås, I., Cappelen, A. W., Sørensen, E., Ø. & Tungodden, B. Fairness and the development of inequality acceptance. *Science* **328**, 1176–1178 (2010).
27. Boehm, C. Impact of the human egalitarian syndrome on Darwinian selection mechanics. *Am. Nat.* **150**, S100–S121 (1997).
28. Kurzban, R. & Leary, M. R. Evolutionary origins of stigmatization: the functions of social exclusion. *Psychol. Bull.* **127**, 187–208 (2001).
29. Hirata, S. & Fuwa, K. Chimpanzees (*Pan troglodytes*) learn to act with other individuals in a cooperative task. *Primates* **48**, 13–21 (2006).
30. Melis, A. P., Hare, B. & Tomasello, M. Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Anim. Behav.* **72**, 275–286 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank all the children and day-care centres involved in this work for their cooperation, and we thank J. Dauksch, G. Duden, A. Gampe, I. Gavriliu, L. Hering, C. Markmann, M. Plötner, E. Rossi, A. Schrimpf, C. Weiske and the practical students at the Wolfgang Köhler Primate Research Center, Leipzig, for their help in recruiting children, data collection and behaviour coding. We also thank R. Pieszek and M. Ulrich for building the apparatuses, S. Tüpke and M. Sureck for creating the figures, R. Mundry for statistical advice and the zookeepers at Leipzig Zoo for their assistance with the chimpanzees. This work received funding from the EC Seventh Framework Program (FP7/2007-2013) under grant agreement no. 215805.

**Author Contributions** K.H., F.W., J.R.G. and M.T. designed the study and wrote the paper. K.H. and J.R.G. conducted the studies and analysed the data. F.W. supervised the research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.H. ([khamann@eva.mpg.de](mailto:khamann@eva.mpg.de)) or F.W. ([warneken@wjh.harvard.edu](mailto:warneken@wjh.harvard.edu)).

## CORRIGENDUM

doi:10.1038/nature10281

### **Migrastatin analogues target fascin to block tumour metastasis**

Lin Chen, Shengyu Yang, Jean Jakoncic, J. Jillian Zhang  
& Xin-Yun Huang

*Nature* **464**, 1062–1066 (2010)

In this Letter, we reported the crystal structure of macroketone bound to fascin (Protein Data Bank number PDB 3LNA) (Fig. 2d). The chemical structure of macroketone was incorrectly shown and has been corrected (PDB 308K). We have been advised that the crystallographic data for this complex are not technically robust, and do not justify the conclusion that macroketone is bound as shown in Fig. 2d. We therefore regretfully withdraw the X-ray structure models PDB 3LNA and PDB 308K from this study. However, we believe that the rest of the Letter, including the observations made using the mutants, is not directly affected.

## CORRIGENDUM

doi:10.1038/nature10294

### **Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA**

Rong-guang Zhang, Katherine M. Pappas, Jennifer L. Brace,  
Paula C. Miller, Tim Oulmassov, John M. Molyneaux,  
John C. Anderson, James K. Bashkin, Stephen C. Winans  
& Andrzej Joachimiak

*Nature* **417**, 971–974 (2002)

In this Letter, the name of author Terina Pappas should be Katherine M. Pappas. This has been corrected in the HTML version.



# Decoherence in crystals of quantum molecular magnets

S. Takahashi<sup>1,2,3</sup>, I. S. Tupitsyn<sup>4,5</sup>, J. van Tol<sup>6</sup>, C. C. Beedle<sup>7†</sup>, D. N. Hendrickson<sup>7</sup> & P. C. E. Stamp<sup>4,5</sup>

Quantum decoherence is a central concept in physics. Applications such as quantum information processing depend on understanding it; there are even fundamental theories proposed that go beyond quantum mechanics<sup>1–3</sup>, in which the breakdown of quantum theory would appear as an ‘intrinsic’ decoherence, mimicking the more familiar environmental decoherence processes<sup>4</sup>. Such applications cannot be optimized, and such theories cannot be tested, until we have a firm handle on ordinary environmental decoherence processes. Here we show that the theory for insulating electronic spin systems can make accurate and testable predictions for environmental decoherence in molecular-based quantum magnets<sup>5</sup>. Experiments on molecular magnets have successfully demonstrated quantum-coherent phenomena<sup>6–8</sup> but the decoherence processes that ultimately limit such behaviour were not well constrained. For molecular magnets, theory predicts three principal contributions to environmental decoherence: from phonons, from nuclear spins and from intermolecular dipolar interactions. We use high magnetic fields on single crystals of Fe<sub>8</sub> molecular magnets (in which the Fe ions are surrounded by organic ligands) to suppress dipolar and nuclear-spin decoherence. In these high-field experiments, we find that the decoherence time varies strongly as a function of temperature and magnetic field. The theoretical predictions are fully verified experimentally, and there are no other visible decoherence sources. In these high fields, we obtain a maximum decoherence quality-factor of  $1.49 \times 10^6$ ; our investigation suggests that the environmental decoherence time can be extended up to about 500 microseconds, with a decoherence quality factor of  $\sim 6 \times 10^7$ , by optimizing the temperature, magnetic field and nuclear isotopic concentrations.

Environmental decoherence processes are reasonably well understood at the atomic scale<sup>9</sup> (although some poorly understood noisy sources remain<sup>10</sup>). However both quantum information processing, and the fundamental tests noted above, require an understanding of decoherence in larger systems, where experimental decoherence rates are usually much larger than theoretical predictions. This discrepancy is usually attributed to ‘extrinsic’ sources (external noise, uncontrolled disorder/impurities). We thus need to find systems, with many degrees of freedom, where extrinsic decoherence can be eliminated, and where we have a quantitative understanding of other decoherence sources.

Many insulating electronic spin systems are currently the subject of intense experimental interest, notably in semiconductor quantum dots<sup>11,12</sup>, nitrogen-vacancy centres in diamond<sup>13–15</sup> and large-spin magnetic molecules<sup>6–8</sup>. In all these systems, three environmental decoherence mechanisms are involved. The electronic spins couple locally to (1) phonons (an oscillator bath<sup>16</sup>); (2) to large numbers of nuclear spins (a spin bath<sup>17</sup>); and (3) to each other via dipolar interactions. The long range of dipolar interactions is a major problem: it makes quantum error correction more difficult, is theoretically complicated<sup>18</sup> and is very hard to eliminate experimentally.

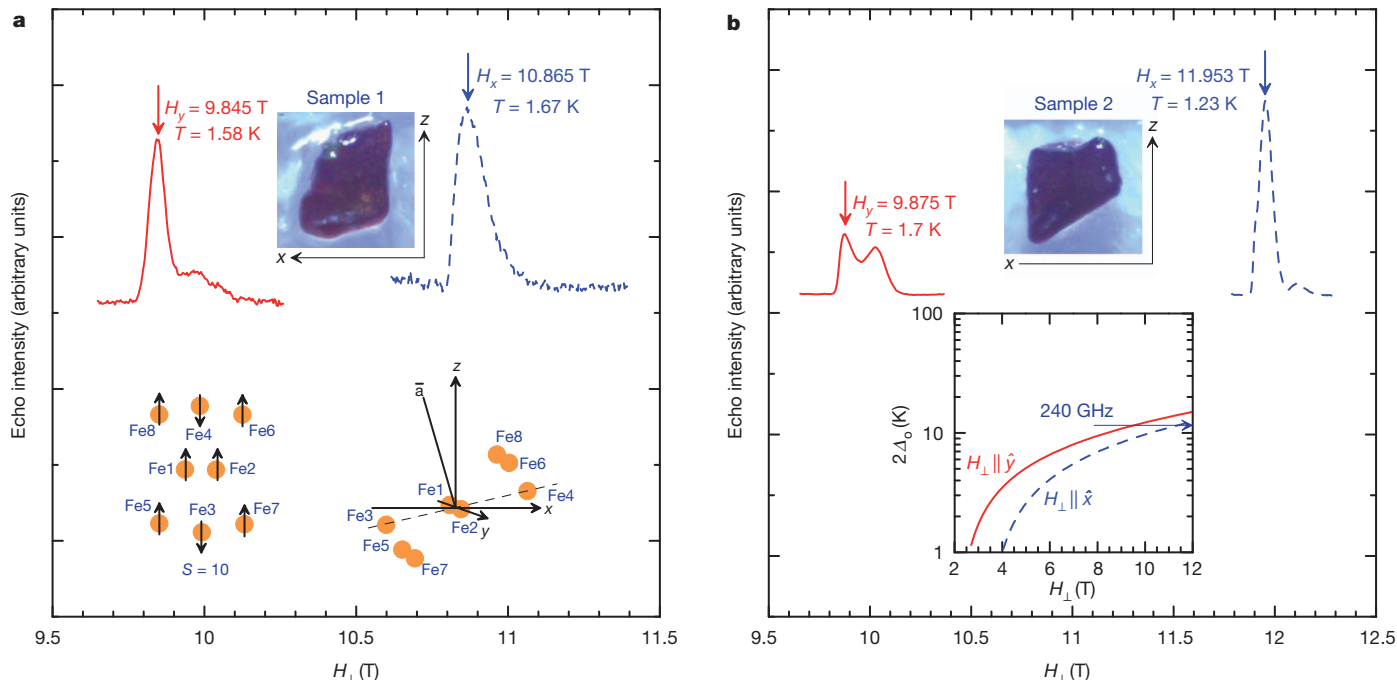
In these crystalline Fe<sub>8</sub> molecular magnets, the electronic spins are structurally ordered, and quantum coherence is observed in the collective (magnon) motion of the spins, rather than in single qubit dynamics. Two great advantages of the Fe<sub>8</sub> system<sup>19</sup> are that the interaction strengths are well known, allowing quantitative predictions, and that it can be prepared with little disorder and few impurities, reducing the danger of extrinsic decoherence. The number of relevant environmental degrees of freedom is very large; depending on isotopic concentrations, there are  $10^{50}$ – $10^{54}$  nuclear spin levels in each molecule, and the system couples to a bulk phonon bath.

In the spin echo experiments described here, a Hahn echo sequence<sup>20</sup> was created in two single crystals of Fe<sub>8</sub> molecules, with natural isotopic concentrations, using a 240-GHz pulsed ESR (electron spin resonance) spectrometer<sup>21,22</sup>. Thus a uniform ESR precession mode (a  $\mathbf{k} = 0$  magnon, where  $\mathbf{k}$  is the wave vector of the magnon) interacts with its surroundings, and we measure the decoherence time  $T_2$ . At low temperature, each electronic spin system behaves as a two-level quantum bit (qubit), with a splitting  $2\Delta_0$  that depends strongly on the local transverse field  $\mathbf{H}_\perp$ , perpendicular to the easy axis  $\hat{z}$  (see Fig. 1 inset). Almost all previous experiments on electron spin systems examined the low-field regime, where nuclear spin decoherence is very strong; here we go to the high-field regime, where its effects are much weaker. Typical ESR results are shown in Fig. 1; we discuss them in detail below.

To understand  $T_2$ , and the ESR lineshape, we need to look at the processes contributing to them. For convenience, we define a dimensionless decoherence rate  $\gamma_\phi = \hbar/\Delta_0 T_2$  (where  $\hbar$  is Planck’s constant  $h$  divided by  $2\pi$ ) and the associated ‘decoherence Q-factor’,  $Q_\phi = \pi/\gamma_\phi$ . Then the processes contributing to  $\gamma_\phi$  are as follows (the full quantitative discussion, for the two samples in this experiment, is given in Supplementary Information):

First, nuclear spins interact locally with each molecular spin, and cause decoherence by a ‘motional narrowing’ process in which they attempt to entangle with the fast-moving qubit<sup>18,23</sup>. The nuclear decoherence rate is  $\gamma_\phi^N = E_0^2/2\Delta_0^2$ , where  $E_0$  is the half-width of the Gaussian multiplet of nuclear spin states coupled to the qubit; and the nuclear contribution to the ESR linewidth is just  $E_0$ . Now in this experiment, with naturally occurring isotopic concentrations,  $E_0 = 4 \times 10^{-4}$  K at these fields, where  $\Delta_0 = 5.75$  K. Thus  $\gamma_\phi^N \approx 10^{-9}$  is very small, simply because  $\Delta_0$  is so large in these fields; and the nuclear spin contribution ( $\sim E_0$ ) to the linewidth is also very small compared to the main contributions. Isotopic substitution of deuterium for the 120 protons in each molecule will further decrease  $\gamma_\phi^N$  by a factor of 15.2 to  $\gamma_\phi^N \approx 7 \times 10^{-11}$ . In principle, there can be a ‘noise’ contribution from the intrinsic nuclear dynamics caused by internuclear interactions<sup>17</sup>; however, in contrast to quantum dot systems<sup>24</sup>, such contributions are very small in molecular magnets<sup>23</sup> (even in systems with strongly interacting nuclei like Mn<sub>12</sub>, where they have been seen<sup>25</sup>).

<sup>1</sup>Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA. <sup>2</sup>Institute for Terahertz Science and Technology, University of California, Santa Barbara, California 93106, USA. <sup>3</sup>Department of Physics, University of California, Santa Barbara, California 93106, USA. <sup>4</sup>Pacific Institute of Theoretical Physics, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. <sup>5</sup>Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. <sup>6</sup>National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, USA. <sup>7</sup>Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA. <sup>†</sup>Present address: National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, USA.



**Figure 1 | Typical ESR spectra, showing echo intensity as a function of transverse magnetic field,  $H_{\perp}$ .** Data are shown for two different samples, at different temperatures and orientations of field, and at  $\omega_{\text{ESR}} = 240$  GHz. **a**, Sample 1. Solid red line,  $H_{\perp} \parallel \hat{y}$ ,  $T = 1.58$  K; dashed blue line,  $H_{\perp} \parallel \hat{x}$ ,  $T = 1.67$  K. Top inset, sample dimensions are approximately  $z: x: y = 1,000: 700: 250$   $\mu\text{m}$ . Lower left inset, the low- $T$  spin structure of the  $\text{Fe}_8$  molecule.

Second, the form of the local spin–phonon interaction is determined by the system symmetry. At high fields this interaction simplifies, and we find a dimensionless phonon decoherence rate<sup>18</sup> given by  $\gamma_{\phi}^{\text{ph}} = [(\mathcal{F}_{AS} \Delta_o^2) / (\pi \rho c_s^3 \hbar^3)] \coth(\Delta_o / k_B T)$ , where  $\rho$  is the sample density,  $c_s$  the sound velocity, and  $\mathcal{F}_{AS}$  the relevant spin–phonon matrix element. The contribution of this spin–phonon process to the ESR linewidth is negligible.

Third, the intermolecular dipole interaction directly couples the  $\mathbf{k} = 0$  ESR precession mode to finite-momentum magnons; it may decay spontaneously into multiple magnons, or scatter off existing thermal magnons. This process affects the ESR lineshape and the decoherence rate very differently. The long-range dipolar interaction creates a distribution of demagnetization fields around the sample. In highly polarized samples, this is strongly sample-shape dependent, but for annealed samples, it is Gaussian distributed<sup>26,27</sup>; in both cases it can be calculated numerically. The lineshape then reflects the quite broad distribution of these fields. However, decoherence comes from the magnon decay process described above, and depends only on the phase space available for these processes at the resonance field; it can then be calculated directly from the analytic expression for this process. At the experimental temperatures, the magnon decoherence rate is  $\sim \exp(-2\Delta_o / k_B T)$ , where  $k_B$  is Boltzmann’s constant and  $T$  is temperature, coming almost entirely from thermal magnon scattering.

Last, there can be extrinsic contributions from impurities and defects, which typically cause the easy-axis anisotropy parameter  $D$  of the  $\text{Fe}_8$  Hamiltonian to fluctuate around the sample (the ‘ $D$ -strain’ effect<sup>28,29</sup>). This will then contribute to the ESR linewidth. Such static impurities and defects can cause  $\Delta_o$  to vary in the sample (although we find no evidence for such a spread in this experiment). However they can not contribute to decoherence at all, provided they are static, because then they simply shift the individual qubit energies. On the other hand any impurities or defects with significant dynamics will cause extrinsic decoherence.

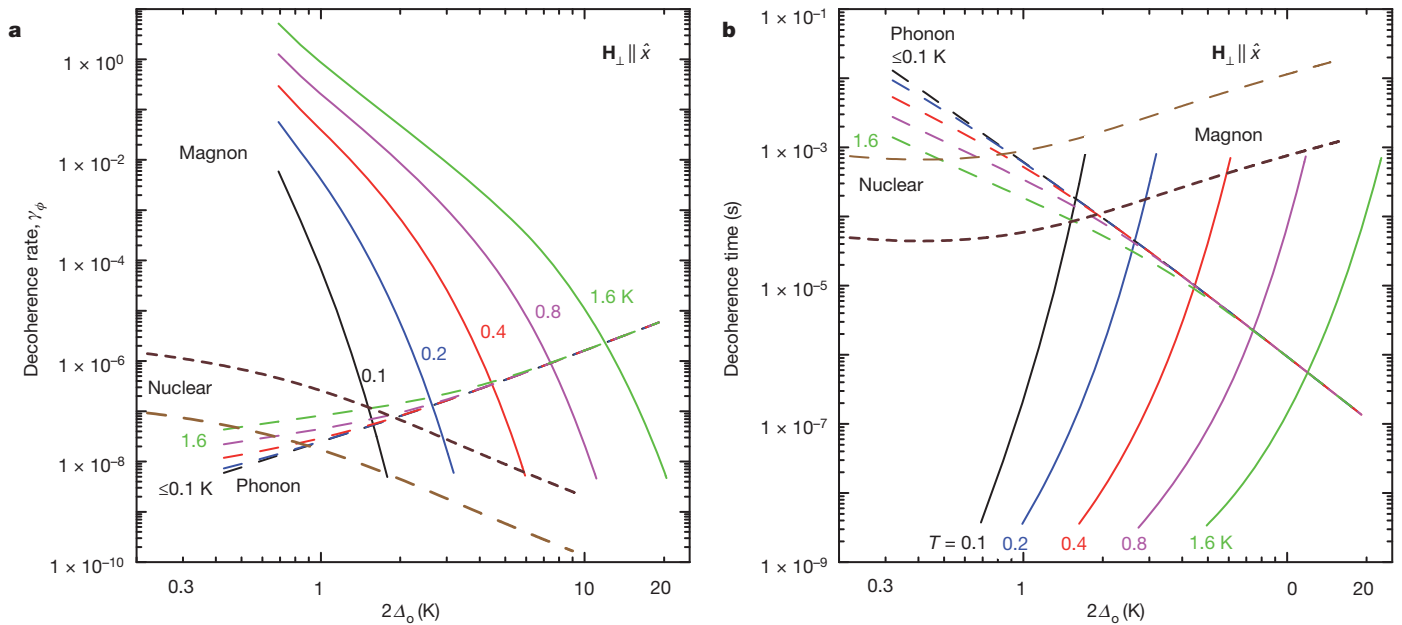
In Fig. 2 we show how these contributions to the decoherence rate are predicted to vary with field and temperature for a crystalline  $\text{Fe}_8$

system. The spin–phonon contribution  $\gamma_{\phi}^{\text{ph}}$  increases with applied transverse field, because the available phonon phase space increases sharply with  $\Delta_o$ . However the nuclear spin decoherence rate decreases with field: roughly  $\gamma_{\phi}^{\text{N}} \propto 1/\Delta_o^2$ , because  $E_o$  changes quite slowly with field (and also decreases as  $\Delta_o$  increases). Thus there is a crossover, with a minimum  $\gamma_{\phi} \approx 10^{-7}$  when  $\Delta_o \approx 1$  K. However these two ‘single-qubit’ decoherence mechanisms are entirely masked for a dense crystal by the dipolar ‘magnon’ decoherence, except at high fields (where dipolar decoherence competes with phonon decoherence) or at very low temperatures and low fields (where it competes with nuclear spin decoherence). In this experiment, we chose to go to high fields.

With all this in mind, we return to the ESR results obtained by echo-detected field sweep, in Fig. 1. The resonant peaks are broadened, with a width  $\sim 0.1$  T; the peculiar structure of the peak when  $H_{\perp} \parallel \hat{y}$ , discussed in detail in Supplementary Information, comes from dipolar interactions. These ESR signals may be understood as follows. The qubit splitting  $2\Delta_o$  varies with field as shown in Fig. 1b inset. For fields  $H_y = 9.5$  T,  $H_x = 0$ , or  $H_x = 11.3$  T,  $H_y = 0$ , the electronic spin Hamiltonian<sup>18,30</sup> for  $\text{Fe}_8$  predicts  $2\Delta_o(H_{\perp}) \approx 11.5$  K, equivalent to our spectrometer frequency of 240 GHz (see Fig. 1 inset), implying we should see resonance peaks at these fields. These predictions are reasonably well satisfied in both samples. The discrepancies, discussed in detail in Supplementary Information, come from two sources: (1) sample misorientation, and (2) weak departures at high field from the model Hamiltonian<sup>18,30</sup> used to predict the field splitting.

The results of the measurements for each sample, together with the calculated theoretical decoherence times for  $\text{Fe}_8$ , are presented in Fig. 3. The agreement is very good; we emphasize that apart from the size of the spin–phonon coupling, which is not known exactly, there are no adjustable parameters in these fits. The decoherence times and rates in the experiment range over roughly an order of magnitude, with a maximum  $T_2 \approx 0.63$   $\mu\text{s}$ , corresponding to  $Q_{\phi} \approx 1.49 \times 10^6$ , at the lowest temperature we went to.

A number of features should be stressed here. First, notice how differently the decoherence and the ESR lineshape are affected by



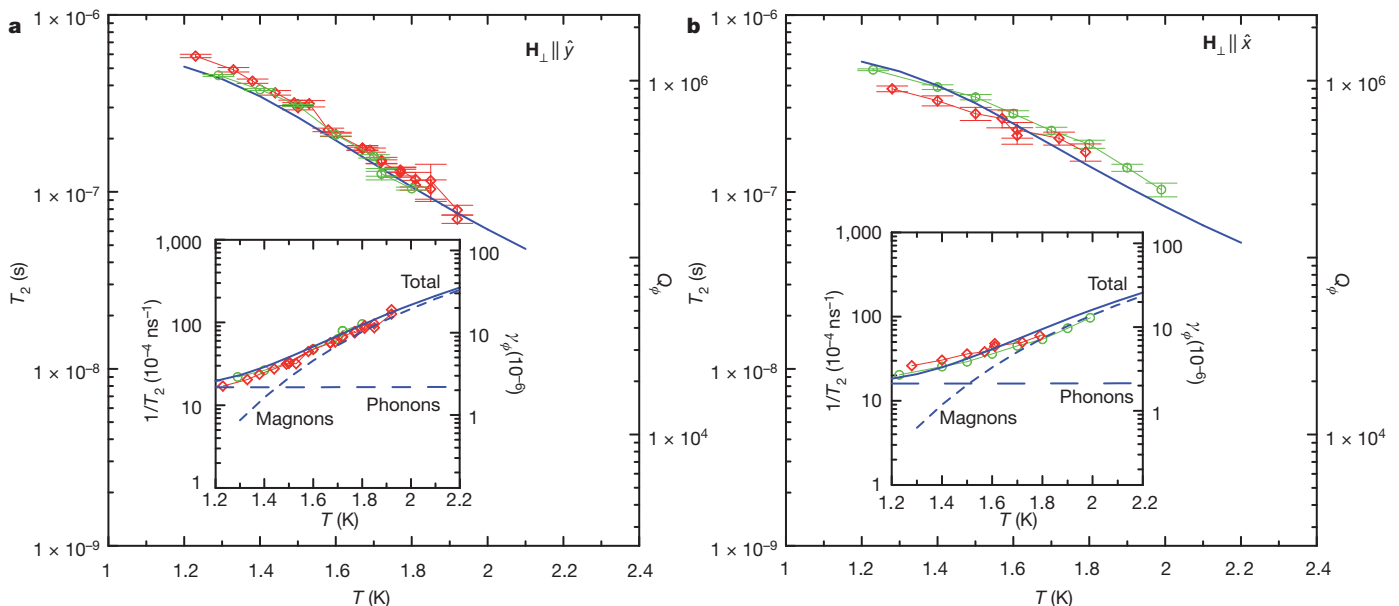
**Figure 2 | Calculated contributions to the decoherence coming from the coupling to nuclear spins, phonons and magnons.** **a**, The three individual contributions which sum to give the dimensionless decoherence rate  $\gamma_\phi = \hbar/T_2\Delta_0$ , as a function of the qubit splitting in the case  $\mathbf{H}_\perp \parallel \hat{x}$ . **b**, The three corresponding contributions to the decoherence time,  $T_2$ . In both panels:

the different environmental couplings. The ESR linewidth and line-shape are completely dominated by static impurity fields and by the spatially varying dipolar fields. However the decoherence is completely dominated, at these high fields, by the phonons and dipolar interactions. At lower fields, the nuclear spin decoherence would also be important, but its effect on the ESR linewidth would still be negligible. Note also that whereas the dipolar contribution to the ESR lineshape depends strongly on sample shape, this shape can only affect the

brown dashed lines, nuclear contribution (short-dashed brown lines are for the natural isotopic concentrations, long-dashed brown lines for the deuterated system); solid lines of different colours, magnon contributions at different temperatures (shown) from 0.1 to 1.6 K; long-dashed lines of different colours, phonon contributions, shown for the same temperatures as in the magnon case.

dipolar decoherence near the edges of this line. In the middle of the line, when the decoherence is coming from molecules near the centre of the sample where the field is homogeneous, one expects no dependence of the dipolar decoherence on sample shape. This is also what we found in the experiment.

Second, we emphasize how the experiment tests the phonon and dipolar contributions to the decoherence separately: they have very different temperature dependences in the regime covered here, with



**Figure 3 | Measured and calculated decoherence times  $T_2$  in samples 1 and 2, as a function of temperature.** **a**, Results for  $\mathbf{H}_\perp \parallel \hat{y}$ . Main panel: thin red line with diamonds, measured using sample 1,  $H_y = 9.845$  T; thin green line with circles, measured using sample 2,  $H_y = 9.875$  T; vertical and horizontal error bars, standard errors of  $T_2$  data fits and uncertainty in temperature ( $\Delta T = \pm 0.05$  K), respectively. Thick blue line, calculations including phonon and magnon contributions,  $H_y = 9.5$  T. Inset: partial contributions calculated

for  $T_2^{-1}$  (solid line) from magnons (dashed line) and phonons (long-dashed line), together with the corresponding experimental results for the two samples (diamonds and circles). The scale on the right-hand side of the main panel indicates the decoherence  $Q$ -factor,  $Q_\phi = \pi/\gamma_\phi = \pi T_2 \Delta_0/\hbar$ ; the right-hand scale on the inset shows  $\gamma_\phi$ . **b**, As for **a**, but now for  $\mathbf{H}_\perp \parallel \hat{x}$ . The experimental curves were measured at  $H_x = 10.865$  T (sample 1) and  $H_x = 11.953$  T (sample 2). The theoretical curves are obtained at  $H_x = 11.3$  T.



phonons dominating below  $\sim 1.2$  K, and magnons dominating at higher temperature. We find agreement between theory and experiment, with no adjustable parameters, across roughly an order of magnitude in decoherence rate. Thus all decoherence in the experiment can be accounted for by environmental sources. This implies that we have no measurable extrinsic decoherence here, either from disorder or dynamic impurities. Nor do we have evidence for any other contributions, either from “third-party decoherence”<sup>4</sup>, or from any of the ‘intrinsic decoherence’ sources<sup>1</sup> discussed in the literature.

Third, we note a key difference between the coherence here, which involves a macroscopic number of qubits excited coherently in a spin wave, and that in other qubit systems, in which entanglement is achieved locally, involving just one or a few qubits. The reason we can do this is because we are dealing with a single crystal.

Last, the present investigation suggests that one can optimize the qubit decoherence  $T_2$  and Q-factor  $Q_\phi$ , as a function of field and temperature, using the results plotted in Fig. 2. We see that lower temperature allows use of a smaller ESR frequency  $2A_0$ ; these two changes strongly reduce the dipolar and phonon decoherence contributions, giving a large increase in  $T_2$ , and a somewhat smaller increase in  $Q_\phi$ . The optimal decoherence rate comes when the phonon and nuclear spin decoherence contributions cross (for natural isotopic concentrations, this is at  $2A_0 \approx 2$  K), provided also that  $T < 0.13$  K, so that the dipolar/magnon decoherence can be ignored. One then finds that  $\gamma_\phi \approx 1.5 \times 10^{-7}$ , so that  $Q_\phi = 2 \times 10^7$ , corresponding to a decoherence time  $T_2 \approx 50 \mu\text{s}$ . However with isotopic substitution of deuterons in place of the protons, the optimal decoherence time rises to  $T_2 \approx 500 \mu\text{s}$ , at  $2A_0 = 0.8$  K  $\equiv 17$  GHz, and  $T = 45$  mK. This corresponds to  $\gamma_\phi \approx 5 \times 10^{-8}$  and  $Q_\phi \approx 6 \times 10^7$ . These considerations show the usefulness of this kind of theory in the optimal design of spin qubit systems—notice the crucial importance of controlling the dipolar interactions between qubits. Notice also that if quantum mechanics is to be tested on anything but microscopic scales, it will be essential to continue developing theory and experiment for systems like the present one, where the environmental decoherence processes can be understood quantitatively, and where extrinsic decoherence sources can be largely eliminated.

## METHODS SUMMARY

The single molecule characteristics were calculated using previous results for crystal field parameters<sup>30</sup> and for the effect of high field on these<sup>18</sup>. Hyperfine couplings were taken from previous work<sup>18</sup>; the spin-phonon couplings were estimated using standard magnetostriction theory, in the high field regime. The dipolar fields were calculated numerically, taking into account the unit cell crystal structure, the sample shape and the field direction, for each sample. The nuclear spin and phonon decoherence rates were determined analytically, using standard methods<sup>16,17</sup>, using the previously determined hyperfine and spin-phonon couplings. The dipolar decoherence rate, from four-magnon scattering and decay processes, was determined numerically, for the given sample shape, field and crystal lattice structure, using analytic formulas<sup>18</sup> for the magnon spectrum and dipolar coupling functions.

Single crystals of  $\text{Fe}_8$  magnetic molecules were synthesized using the method of ref. 19. Each crystal was indexed and unit-cell parameters were checked to ensure consistency. Continuous-wave/pulsed ESR measurements were carried out using the 240-GHz ESR spectrometer at the National High Magnetic Field Laboratory (NHMFL)<sup>21,22</sup>. The system consists of a 12.5-T superconducting magnet, a 40-mW 240-GHz source, quasi-optics, a superheterodyne detection system, and a  $^4\text{He}$  flow cryostat. The spin decoherence time was measured by a Hahn echo sequence ( $\pi/2$ - $\tau$ - $\pi$ - $\tau$ -echo) where the delay  $\tau$  is varied<sup>20</sup>. The magnetic component of the 240-GHz pulses was perpendicular to the d.c. magnetic field, to generate coherent magnons in the sample; their duration was adjusted to maximize the echo signals, and was typically 200–300 ns.

Received 3 April; accepted 16 June 2011.

Published online 20 July 2011.

1. 't Hooft, G. Quantum gravity as a dissipative deterministic system. *Class. Quantum Gravity* **16**, 3263–3279 (1999).

2. Penrose, R. On gravity's role in quantum state reduction. *Gen. Relativ. Gravit.* **28**, 581–600 (1996).
3. Leggett, A. J. Testing the limits of quantum mechanics: motivation, state of play, prospects. *J. Phys. Condens. Matter* **14**, R415–R451 (2002).
4. Stamp, P. C. E. The decoherence puzzle. *Stud. Hist. Phil. Mod. Phys.* **37**, 467–497 (2006).
5. Gatteschi, D., Sessoli, R. & Villain, J. *Molecular Nanomagnets* (Oxford Univ. Press, 2006).
6. Takahashi, S. *et al.* Coherent manipulation and decoherence of  $S = 10$  single-molecule magnets. *Phys. Rev. Lett.* **102**, 087603 (2009).
7. Bertina, S. *et al.* Quantum oscillations in a molecular magnet. *Nature* **453**, 203–206 (2008).
8. Schlegel, C. *et al.* Direct observation of quantum coherence in single-molecule magnets. *Phys. Rev. Lett.* **101**, 147203 (2008).
9. Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
10. Labaziewicz, J. *et al.* Temperature dependence of electric field noise above gold surfaces. *Phys. Rev. Lett.* **101**, 180602 (2008).
11. Xu, X. *et al.* Optically controlled locking of the nuclear field via coherent dark-state spectroscopy. *Nature* **459**, 1105–1109 (2009).
12. Ladd, T. D. *et al.* Pulsed nuclear pumping and spin diffusion in a single charged quantum dot. *Phys. Rev. Lett.* **105**, 107401 (2010).
13. Takahashi, S. *et al.* Quenching spin decoherence in diamond through spin bath polarization. *Phys. Rev. Lett.* **101**, 047601 (2008).
14. Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
15. Hanson, R. *et al.* Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science* **320**, 352–355 (2008).
16. Leggett, A. J. *et al.* Dynamics of the dissipative two-state system. *Rev. Mod. Phys.* **59**, 1–85 (1987).
17. Prokof'ev, N. V. & Stamp, P. C. E. Theory of the spin bath. *Rep. Prog. Phys.* **63**, 669–726 (2000).
18. Morello, A., Stamp, P. C. E. & Tupitsyn, I. S. Pairwise decoherence in coupled spin qubit networks. *Phys. Rev. Lett.* **97**, 207206 (2006).
19. Wiegand, K., Pohl, K., Jibril, I. & Huttner, G. Hydrolysis products of the monomeric amine complex  $(\text{C}_6\text{H}_{15}\text{N}_3)\text{FeCl}_3$ : the structure of the octameric iron(III) cation of  $[(\text{C}_6\text{H}_{15}\text{N}_3)_6\text{Fe}_8(\mu_3\text{-O})_2(\mu_2\text{-OH})_{12}]\text{Br}_7 \cdot 8\text{H}_2\text{O}$ . *Angew. Chem. Int. Edn Engl.* **23**, 77–78 (1984).
20. Hahn, E. L. Spin echoes. *Phys. Rev.* **80**, 580–594 (1950).
21. van Tol, J., Brunel, L. C. & Wylde, R. J. A quasi-optical transient electron spin resonance spectrometer operating at 120 and 240 GHz. *Rev. Sci. Instrum.* **76**, 074101 (2005).
22. Morley, G. W., Brunel, L. C. & van Tol, J. A. Multifrequency high-field pulsed electron paramagnetic resonance/electron-nuclear double resonance spectrometer. *Rev. Sci. Instrum.* **79**, 064703 (2008).
23. Stamp, P. C. E. & Tupitsyn, I. S. Coherence window in the dynamics of quantum nanomagnets. *Phys. Rev. B* **69**, 014401 (2004).
24. Yao, W., Liu, R. B. & Sham, L. J. Theory of electron spin decoherence by interacting spins in a quantum dot. *Phys. Rev. B* **74**, 195301 (2006).
25. Morello, A., Bakharev, O. N., Brom, H. B., Sessoli, R. & de Jongh, L. J. Nuclear Spin Dynamics in the Quantum regime of a single molecule magnet. *Phys. Rev. Lett.* **93**, 197202 (2004).
26. Cuccoli, A. *et al.* Dipolar interaction and incoherent quantum tunneling: a Monte Carlo study of magnetic relaxation. *Europhys. J. B* **12**, 39–46 (1999).
27. Tupitsyn, I. S., Stamp, P. C. E. & Prokof'ev, N. V. Hole digging in ensembles of tunneling molecular magnets. *Phys. Rev. B* **69**, 132406 (2004).
28. Hill, S. *et al.* D-strain, g-strain, and dipolar interactions in the  $\text{Fe}_8$  and  $\text{Mn}_{12}$  single molecule magnets: an EPR lineshape analysis. *Int. J. Mod. Phys. B* **16**, 3326–3329 (2002).
29. Park, K. *et al.* Effects of D-strain, g-strain, and dipolar interactions on EPR linewidths of the molecular magnets  $\text{Fe}_8$  and  $\text{Mn}_{12}$ . *Phys. Rev. B* **65**, 014426 (2001).
30. Barra, A. L., Gatteschi, D. & Sessoli, R. High-frequency EPR spectra of  $[\text{Fe}_8\text{O}_2(\text{OH})_{12}(\text{tacn})_6]\text{Br}_8$ : a critical appraisal of the barrier for the reorientation of the magnetization in single-molecule magnets. *Chem. Eur. J.* **6**, 1608–1614 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the NSF (DMR-0520481, DMR-0703925), the Keck Foundation (S.T. and J.v.T.), NSERC, CIFAR, PITP, the John E. Fetzer Memorial Trust (grant D21-C62) and the Center for Philosophy and the Natural Sciences, California State University, Sacramento (I.S.T. and P.C.E.S.). The National High Magnetic Field Laboratory is supported by NSF Cooperative Agreement DMR-0654118, by the State of Florida, and by the DOE.

**Author Contributions** S.T., I.S.T. and P.C.E.S. contributed to the writing of the manuscript. S.T., I.S.T. and P.C.E.S. conceived the ESR experiments. The ESR measurements were carried out by S.T. and J.v.T. The theoretical work was done by I.S.T. and P.C.E.S. C.C.B. and D.N.H. synthesized  $\text{Fe}_8$  crystals and performed X-ray diffraction measurements.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.T. ([susumuta@usc.edu](mailto:susumuta@usc.edu)) or P.C.E.S. ([stamp@phas.ubc.ca](mailto:stamp@phas.ubc.ca)).

# The landscape of recombination in African Americans

A list of authors and their affiliations appears at the end of the paper

Recombination, together with mutation, gives rise to genetic variation in populations. Here we leverage the recent mixture of people of African and European ancestry in the Americas to build a genetic map measuring the probability of crossing over at each position in the genome, based on about 2.1 million crossovers in 30,000 unrelated African Americans. At intervals of more than three megabases it is nearly identical to a map built in Europeans. At finer scales it differs significantly, and we identify about 2,500 recombination hotspots that are active in people of West African ancestry but nearly inactive in Europeans. The probability of a crossover at these hotspots is almost fully controlled by the alleles an individual carries at *PRDM9* ( $P$  value  $< 10^{-245}$ ). We identify a 17-base-pair DNA sequence motif that is enriched in these hotspots, and is an excellent match to the predicted binding target of *PRDM9* alleles common in West Africans and rare in Europeans. Sites of this motif are predicted to be risk loci for disease-causing genomic rearrangements in individuals carrying these alleles. More generally, this map provides a resource for research in human genetic variation and evolution.

In humans and many other species, recombination is not evenly distributed across the genome, but instead occurs in 'hotspots': 2-kilobase (kb) segments where the crossover rate is far higher than in the flanking DNA sequence<sup>1–3</sup>. The highest-resolution genetic map in contemporary humans so far—the deCODE map—is based on about 500,000 crossovers identified in 15,000 Icelandic meioses<sup>4</sup>. However, a limitation of maps built in people of European descent<sup>4–6</sup> is that they may not apply equally well in other populations, as suggested by comparisons of maps across ethnic groups<sup>4,7–9</sup> and patterns of linkage disequilibrium breakdown, which indicate that more of the genome may be recombinationally active in West Africans<sup>10</sup>. It is known that a major determinant of the positions of recombination hotspots is *PRDM9*, a meiosis-specific histone H3 methyltransferase whose zinc finger (ZF) domain binds DNA sequence motifs<sup>11–13</sup>. In Europeans, *PRDM9* ZF arrays are predominantly of two similar types, A and B, both of which bind the 13-bp motif CCNCCNTNCCNC<sup>11</sup>. In contrast, 36% of West African alleles are not of the A or B type<sup>9,13</sup>. Sperm typing of males who carry neither the A nor the B allele has shown no evidence of crossover activity at recombination hotspots associated with the 13-bp motif<sup>9</sup>.

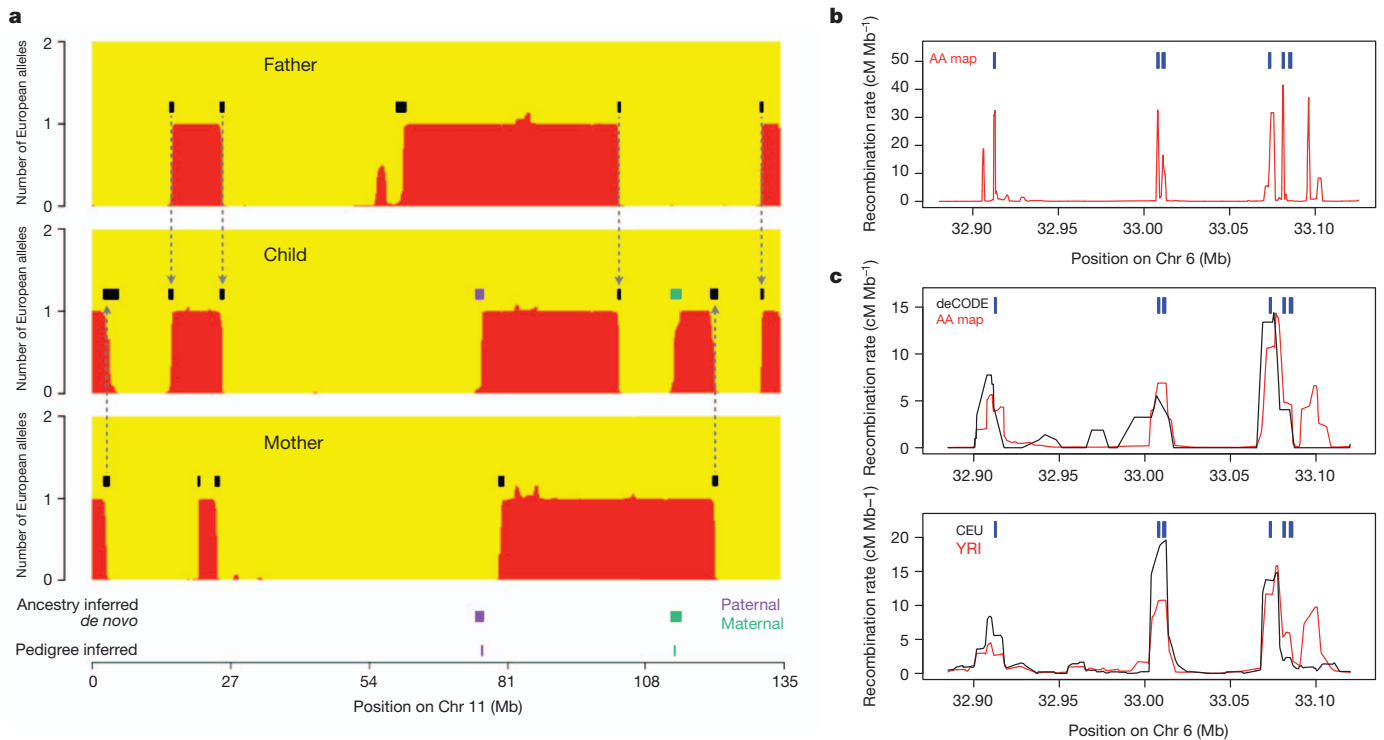
## Building an African–American genetic map

To investigate differences in the crossover landscape across human populations, we built a genetic map in African Americans, who have an average of about 80% West African and 20% European ancestry, leading to genomes comprised of multi-megabase stretches of either West African or European ancestry<sup>14</sup>. Computational approaches, including HAPMIX<sup>15</sup>, have been developed to infer the probability of 0, 1 or 2 European or African alleles at each locus in individuals genotyped at hundreds of thousands of single nucleotide polymorphisms (SNPs)<sup>15–17</sup>. Positions where the inferred number of European or African alleles changes reflect crossover events that have occurred since admixture began (on average six generations ago<sup>15</sup>). Change in the probability of European ancestry between adjacent SNPs can be interpreted as the probability of such a crossover between them. We inferred crossover events in 29,589 apparently unrelated African Americans who had been genotyped on SNP arrays in genetic association studies (Methods; Fig. 1a). To minimize false-positive crossovers, we restricted

to crossovers that HAPMIX inferred with a probability of  $>95\%$ , and that were flanked by a minimum of 2-centimorgan (cM) stretches where the ancestry was inferred to be unchanging (Supplementary Note 1). This produced 2,113,293 high-confidence crossovers, with a typical switch point resolved within 70 kb with probability 50% (Supplementary Note 1).

To build a high-resolution African-American genetic map (AA map), we leveraged the fact that most crossovers occur in hotspots shared across individuals<sup>2</sup> (Methods). Intuitively, although any crossover can only be roughly localized, inter-SNP intervals that are inferred to have an appreciable probability of crossover in multiple individuals are likely to contain recombination hotspots, allowing much better localization (Supplementary Fig. 1). To implement this idea, we modelled the recombination rate for each inter-SNP interval as shared across individuals and used Markov chain Monte Carlo (MCMC) to sample rates consistent with the data (Methods). This provides well-calibrated estimates of the crossing-over rate between all pairs of markers as well as estimates of rate uncertainty (Supplementary Note 1 and Supplementary Fig. 2). We find that the interval size at which the average recombination rate is equal to the standard error is 6 kb, which is the same accuracy that would be expected from a map based on 500,000 crossovers whose boundaries were precisely resolved (Supplementary Note 1). Despite this high resolution, there are also some limitations. First, the AA map does not separately infer male and female recombination rates (it is a sex-averaged map) and requires normalization by the total map length (like linkage disequilibrium maps<sup>3,18</sup>). Second, the map has less resolution and may miss a higher fraction of true crossovers at loci where it is more difficult to detect and resolve crossovers owing to low SNP density or low differentiation between West Africans and Europeans. Third, the map may be biased where ancestry deviates from the average, for example at chromosome 8q24, where the 10% of the people in this study who have prostate cancer have an increased proportion of African ancestry<sup>19</sup>. Fourth, the map assumes that all individuals are unrelated, whereas in fact there is probably some shared ancestry, resulting in multiple counting of some crossovers and an overestimation of map precision.

To assess the accuracy of the AA map, we generated an independent African-American pedigree map by analysing 222 nuclear families



**Figure 1 | Building an African-American genetic map. a**, HAPMIX detection of crossovers between segments of inferred ancestry is illustrated in a father-mother-child trio. Black segments show inferred crossovers; arrows show transmission of ancestral crossovers from parent to child; purple/green segments show *de novo* events (paternal/maternal origin, respectively) corresponding to

that included 1,056 meioses in which we could directly detect crossovers between parent and child (Methods; Fig. 1a). Examination of the AA map rate around directly detected crossovers confirms the high resolution: the rate around such crossovers shows at least as strong a peak as that observed in maps based on linkage disequilibrium<sup>2,3,18</sup> (Supplementary Fig. 3). We next computed correlation coefficients for both the AA map and the deCODE map<sup>4</sup> to maps derived from the breakdown of linkage disequilibrium in Europeans (CEU) and West Africans (YRI)<sup>18</sup>. At broad scales (>3 Mb) they are almost identical ( $\rho > 0.97$ ; Table 1). At fine scales, the AA map is more accurate (Table 1 and Supplementary Table 1), as reflected in a modest improvement in correlation to the CEU map at a 3-kb scale ( $\rho_{AA,CEU} = 0.66$  versus  $\rho_{deCODE,CEU} = 0.58$ ), and a major improvement for the YRI map, also at a 3-kb scale ( $\rho_{AA,YRI} = 0.71$  versus  $\rho_{deCODE,YRI} = 0.53$ ). The deCODE map is more correlated to the CEU map than to the YRI map at scales <1 Mb, suggesting that this map, built in Icelanders, reflects more European recombination rates. The AA map shows the opposite pattern, suggesting that it reflects more West African recombination patterns.

**Table 1 | Genetic map assessments at different size scales**

Scale (interval size)	Pearson correlation ( $\rho$ ) of the AA map (deCODE map) to the specified LD map			Estimated correlation of AA map to the true map (inferred by MCMC) <sup>†</sup>	Estimated coefficient of variation of AA map (s.e. divided by crossover rate expected for interval size) <sup>†</sup>
	Combined LD*	CEU	YRI		
3 kb	0.75 (0.63)	0.66 (0.58)	0.71 (0.53)	0.93	1.41
10 kb	0.82 (0.74)	0.73 (0.70)	0.78 (0.65)	0.96	0.73
30 kb	0.86 (0.83)	0.78 (0.78)	0.83 (0.74)	0.98	0.36
100 kb	0.91 (0.89)	0.84 (0.85)	0.87 (0.81)	0.99	0.17
300 kb	0.94 (0.93)	0.89 (0.90)	0.92 (0.88)	1.00	0.08
1 Mb	0.97 (0.96)	0.94 (0.94)	0.95 (0.95)	1.00	0.04
3 Mb	0.98 (0.98)	0.97 (0.97)	0.98 (0.97)	1.00	0.02

The numbers in this table are restricted to the autosomes and genomic segments more than 5 Mb from the telomeres. LD, linkage disequilibrium; s.e., standard error.

\*The combined map is the HapMap2 population-averaged linkage-disequilibrium-based map<sup>18</sup>.

<sup>†</sup>The s.e. of the map at each size scale is determined by the posterior probability distribution from the MCMC.



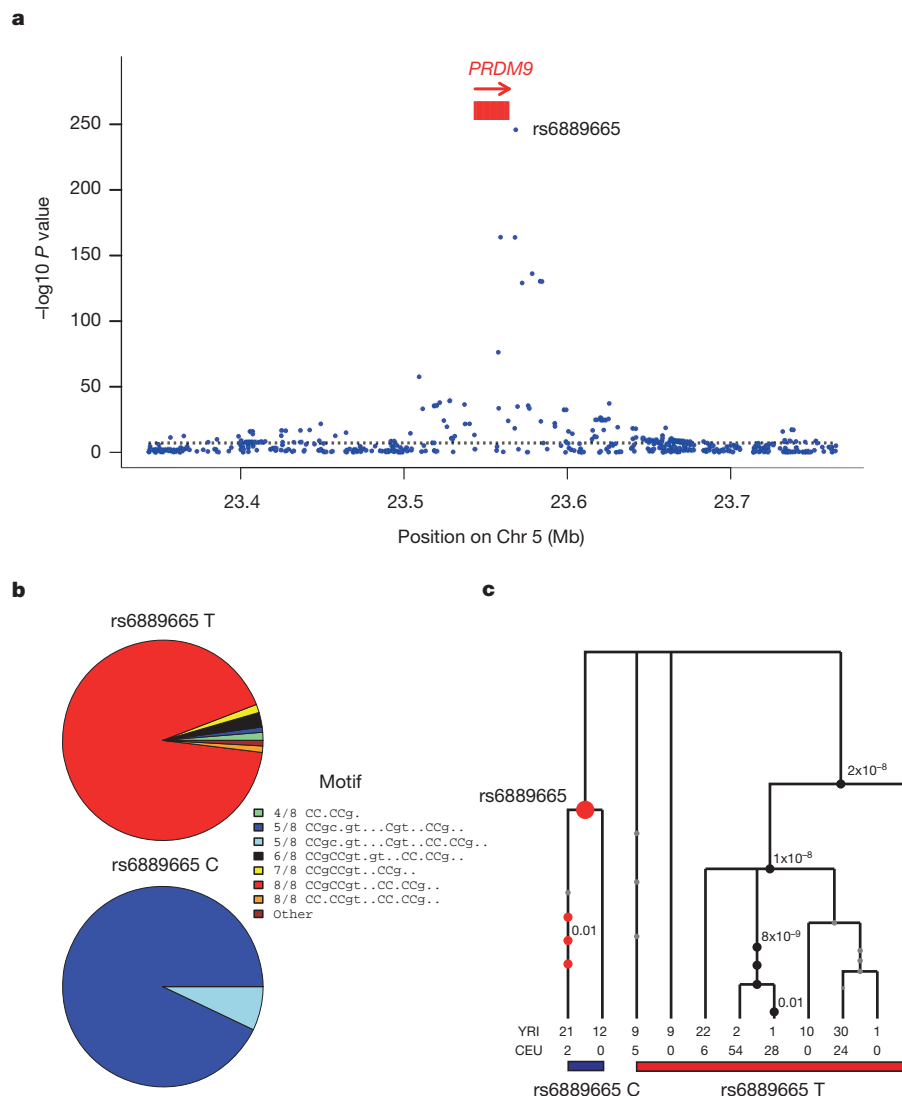
(Methods and Supplementary Fig. 5B). Thus, hotspots active in Europeans are consistently 'shared' with YRI and African Americans, whereas populations with African ancestry harbour additional, non-shared hotspots that we call 'African-enriched'.

### Mapping variants underlying population differences

To understand the features of recombination in West Africans that differ from Europeans, we estimated the degree to which each African-American person's crossovers occur in African-enriched hotspots, compared with shared hotspots, a phenotype we refer to as their African enrichment (AE). We view each individual's crossovers as sampled from a mixture of two genetic maps—an 'S map' of shared hotspots based on the deCODE map, and an 'AE map' of African-enriched hotspots that is learned from comparing the deCODE and AA maps—so that the proportion of crossovers assigned to the AE map is a person's AE phenotype (Supplementary Note 4). We tested approximately 3 million SNPs (genotyped and imputed) for association with three phenotypes: AE, usage of linkage-disequilibrium-based hotspots known to be enriched for the 13-bp motif

CCNCCNTNNCCNC<sup>20</sup> and genome-wide crossover rate (in pedigrees) (Methods and Supplementary Note 4). In crossovers detected in unrelated African Americans, the alleles a person carries are only sometimes descended from the ancestor in whom the crossover occurred, thus adding noise to the association signal (nevertheless there is useful signal given the large sample size; Supplementary Note 4). In the pedigree map, association between alleles and AE can be tested directly because we have genotypes in the parents.

The SNP showing the strongest association with AE is rs6889665 ( $P = 1.5 \times 10^{-246}$ ; Fig. 2a and Supplementary Fig. 6), which has a derived allele frequency of 29% in YRI and 2% in CEU, and is within 4 kb of the ZF array of *PRDM9* (refs 4, 9, 11–13). This SNP is associated with AE in both the pedigree individuals and the unrelated individuals (Supplementary Note 4), and is also the SNP most strongly associated with usage of linkage-disequilibrium-based hotspots ( $P = 1.8 \times 10^{-52}$ ) (Supplementary Table 2). No locus outside *PRDM9* is significant ( $P < 0.01$  after Bonferroni correction; Supplementary Table 2). To understand better the association at rs6889665, we inferred the alleles in the *PRDM9* ZF array carried



**Figure 2 | Association of *PRDM9* genetic variation with hotspot activity.** **a**, A genome-wide association study measuring association of the AE phenotype shows a single genome-wide significant peak at *PRDM9*, with rs6889665 the best-associated SNP. **b**, Relationship between alleles of rs6889665 and predicted binding target of the *PRDM9* ZF array<sup>9</sup> for West African and European samples. The binding predictions are grouped into 8

clusters according to their best-matching region to the 13-bp motif, and annotated by the number of bases matching the motif. The African-enriched rs6889665 C allele always co-occurs with motifs with a poor (5/8) match to the 13-bp motif. **c**, Gene tree<sup>25</sup> of the linkage disequilibrium block containing the *PRDM9* ZF array (Methods); numbered circles show SNPs and significant  $P$  values for association, after conditioning on rs6889665.

by 139 individuals based on sequencing data from the 1000 Genomes Project<sup>10</sup>, using the reads to infer each individual's *PRDM9* alleles among 29 alleles whose full sequences were previously determined<sup>9</sup> (Supplementary Note 5). Grouping *PRDM9* alleles on the basis of how closely their binding target predictions match the 8 non-degenerate bases of the 13-bp motif, following a previously described approach<sup>9</sup>, we find that the ancestral 'T' variant at rs6889665 is strongly correlated to alleles with an exact (8/8) match to the 13-bp motif (including the A and B alleles), whereas the derived 'C' variant is almost perfectly correlated to a group of alleles, all predicted to bind a common, different 17-bp motif—CCgCNgTNNCgtNNCC<sup>9</sup>—which matches the 13-bp motif at only 5 bases (5/8 match; less strongly signalled bases in the motif are in lowercase and 'N' may be any base). This implies a common historical origin for alleles matching this 17-bp motif (Fig. 2b, Supplementary Fig. 7 and Supplementary Note 5). We also experimentally measured the number of ZF domains in *PRDM9* in 354 individuals including 166 African Americans from the pedigree study (Methods). This showed, again, that rs6889665 differentiates *PRDM9* alleles into two different classes, with 96% of haplotypes carrying the ancestral allele having <14 ZFs, and 93% of haplotypes carrying the derived allele having ≥14 ZFs (Supplementary Fig. 7). After conditioning on rs6889665, there is no evidence that ZF array length is associated with the AE phenotype. Several SNPs near the *PRDM9* ZF array show a conditional association signal that is much weaker than rs6889665, but still significant (Fig. 2c, Supplementary Fig. 6 and Supplementary Note 4), with the strongest at rs10043097 ( $P = 8.3 \times 10^{-14}$ ), upstream of the *PRDM9* transcription start site. These SNPs may tag additional variation in the *PRDM9* ZF array, or potentially expression levels.

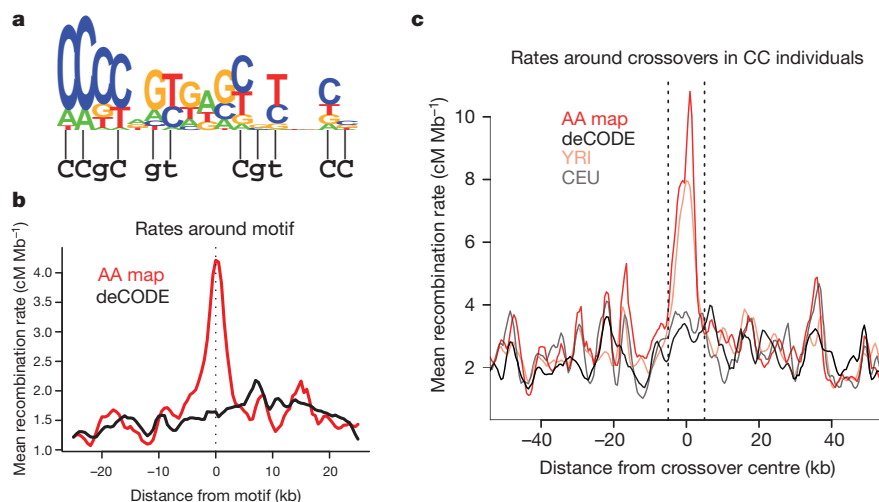
### Finding a motif for African-enriched hotspots

To identify directly candidate African-enriched hotspot motifs, we selected 2,454 loci with a high crossover rate in the AE map and YRI map ( $>2 \text{ cM Mb}^{-1}$  over 2 kb), and no more than half this rate in the S map and CEU map (this set is more powerfully enriched for higher recombination in people of African ancestry than the 2,375 above, as it includes information from the contemporary maps). We compared these to a 'control set' of 7,328 candidate hotspots more active in the European- than the African-derived maps (Methods and Supplementary Note 6). To identify sequence motifs associated with the African-enriched hotspots<sup>3,21</sup>, we identified short motifs that

occurred at increased frequency in the African-enriched hotspot set (Supplementary Note 6). Testing all motifs with lengths of 5–9 bases revealed a 9-nucleotide motif CCCCAGTGA (odds ratio (OR) = 1.79,  $P = 2.24 \times 10^{-8}$ , Bonferroni corrected  $P = 0.004$ ), which exhibited a kilobase-scale rate peak near occurrences of this motif in African-derived maps, but in neither of the European-derived maps (Supplementary Fig. 8). Further analysis revealed a strong influence of downstream flanking bases (Supplementary Fig. 9) and degeneracy, yielding a 17-bp consensus sequence, CCCCAGTGA GCGTgCc (Fig. 3a; more strongly signalled bases are in uppercase), with the same consensus obtained when we considered flanking sequences for only odd or even chromosomes, and whether we based the analysis on AE-S or YRI-CEU map comparisons (Supplementary Note 6). The 500 best matches to this motif have a ~3-fold increase in average rate in the AA and YRI relative to the deCODE and CEU maps (Fig. 3b and Supplementary Fig. 8G). Hotspots associated with the motif occur in both unique and repetitive DNA (for example, L1PA10/13 LINE elements; Supplementary Fig. 10 and Supplementary Note 6). We also compared the 17-bp consensus to the binding motif predicted for 5/8 match alleles, and found that they match almost precisely (Fig. 3a; 10 of 11 bases,  $P = 8.1 \times 10^{-6}$ ).

### Assessing the impact of *PRDM9* on recombination

How much of the African-enriched recombination pattern can be explained by *PRDM9*? We estimated the fraction of variation in the AE phenotype explained by rs6889665 in our pedigree data after accounting for noise in the phenotype estimation (Supplementary Note 4). Over 82% of map usage variability is explained by the rs6889665 genotype alone. Given that there are further influential *PRDM9* variants (Fig. 2c), this gene may thus explain almost all differences in local rate between the West African and European populations. We next examined rates around 82 narrowly defined (<10 kb) crossover sites in 7 individuals homozygous for the derived allele at rs6889665. There is no evidence of hotspots at these loci in either the deCODE or CEU maps (Fig. 3c), in contrast to crossovers in individuals carrying the ancestral allele at rs6889665 (Supplementary Fig. 11). Thus, crossover positions in individuals who are homozygous for the derived allele at rs6889665 are consistent with an entirely different recombination hotspot landscape, which would imply *PRDM9* control of all hotspots<sup>9</sup>. Despite the strong correlation between maps at megabase scales, there is mounting evidence that *PRDM9*'s influence



**Figure 3 | A sequence motif specifying the positions of African-enriched hotspots.** **a**, Logo plot showing a degenerate 17-bp hotspot motif, with stack height proportional to  $-\log P$  value, and relative letter height proportional to the mean crossover rate increase given each base. Below is the bioinformatic *PRDM9* binding prediction for the alleles associated with rs6889665 allele C (from Fig. 2b), matching this motif at 10/11 bases (lines). **b**, Average crossover

rate (in 2-kb sliding windows) in the AA (red line) and deCODE (black line) maps surrounding the 500 strongest motif matches. **c**, In seven rs6889665 CC individuals from the pedigree study, we localized 82 crossovers to within 10 kb, and plot average AA, YRI, deCODE and CEU map rates. There is no strong peak above local background in the deCODE or CEU maps.

on crossing over may not be limited to fine scales<sup>4,11</sup>: we observe a weakly significant association of rs6889665 with the total number of crossovers genome-wide in pedigrees ( $P = 0.04$ ), corresponding to an average 1.3 crossovers more per meiosis per derived allele, exceeding the strongest previously known association<sup>22</sup> at *RNF212*.

## Conclusions

We have shown that *PRDM9* alleles that bind a novel 17-bp motif and occur at greatly increased frequency in people of West African ancestry have led to a shift in the recombination landscape compared with people of non-African ancestry. The larger number of hotspots available to West Africans implies that at the population level, crossovers are more evenly distributed than in Europeans<sup>10</sup>, and thus the shorter extent of West African linkage disequilibrium is not due to differences in demographic history alone (such as the lack of an out-of-Africa founder event)<sup>23</sup>. Our findings also have medical implications, as recombination errors leading to insertions or deletions are known to be associated with recombination hotspots<sup>9,21,24</sup>. Our results predict that the congenital abnormalities that have been associated with the recombination hotspots bound by *PRDM9* A and B alleles will occur at a decreased rate in people of West African ancestry, whereas new diseases will arise due to recombination errors near African-enriched hotspots.

## METHODS SUMMARY

We assembled SNP array data from 29,589 unrelated people and 222 nuclear families genotyped at 490,000–910,000 SNPs from the Candidate Gene Association Resource (CARE), studies at the Children's Hospital of Philadelphia (CHOP), the African American Breast Cancer Consortium, the African American Prostate Cancer Consortium and the African American Lung Cancer Consortium. To build a recombination map, we used HAPMIX to localize candidate crossover positions<sup>15</sup>, and implemented a MCMC that used the probability distributions for the positions of the filtered crossovers to infer recombination rates for each of 1.3 million inter-SNP intervals. We also implemented a second MCMC that models each individual's set of crossovers as a mixture of an S map, similar to the European deCODE map, and an AE map, and then assigned each individual an 'AE phenotype' corresponding to the proportion of their newly detected crossovers assigned to the AE map. We imputed genotypes at up to three million HapMap2 SNPs<sup>18</sup> and then tested each of these SNPs for association with the AE phenotype and other recombination-related phenotypes. We identified 2,454 candidate African-enriched hotspots with increased recombination rates in the YRI versus CEU maps, and in the AE versus S maps, and searched for motifs enriched at these loci, thus identifying a degenerate 17-bp motif. To study the structure of *PRDM9*, we measured the length of the *PRDM9* ZF array and genotyped rs6889665 in YRI, CEU and the CARE nuclear families; we also carried out imputation based on 1000 Genomes Project short read data<sup>10</sup> to infer the alleles individuals carry, among 29 previously characterized in a sequencing study of *PRDM9* (ref. 9).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 February; accepted 27 June 2011.**

**Published online 20 July 2011.**

1. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
2. McVean, G. A. *et al.* The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581–584 (2004).
3. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005).
4. Kong, A. *et al.* Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* **467**, 1099–1103 (2010).
5. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
6. Matisse, T. C. *et al.* A second-generation combined linkage-physical map of the human genome. *Genome Res.* **17**, 1783–1786 (2007).
7. Weitkamp, L. R. Proceedings: population differences in meiotic recombination frequency between loci on chromosome 1. *Cytogenet. Cell Genet.* **13**, 179–182 (1974).
8. Jorgenson, E. *et al.* Ethnicity and human genetic linkage maps. *Am. J. Hum. Genet.* **76**, 276–290 (2005).

9. Berg, I. L. *et al.* *PRDM9* variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genet.* **42**, 859–863 (2010).
10. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
11. Baudat, F. *et al.* *PRDM9* is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
12. Myers, S. *et al.* Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876–879 (2010).
13. Parvanov, E. D., Petkov, P. M. & Paigen, K. *Prdm9* controls activation of mammalian recombination hotspots. *Science* **327**, 835 (2010).
14. Smith, M. W. *et al.* A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* **74**, 1001–1013 (2004).
15. Price, A. L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
16. Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.* **82**, 290–303 (2008).
17. Patterson, N. *et al.* Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000 (2004).
18. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
19. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl Acad. Sci. USA* **103**, 14068–14073 (2006).
20. Coop, G., Wen, X., Ober, C., Pritchard, J. K. & Przeworski, M. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* **319**, 1395–1398 (2008).
21. Myers, S., Freeman, C., Auton, A., Donnelly, P. & McVean, G. A common sequence motif associated with recombination hotspots and genome instability in humans. *Nature Genet.* **40**, 1124–1129 (2008).
22. Kong, A. *et al.* Sequence variants in the *RNF212* gene associate with genome-wide recombination rate. *Science* **319**, 1398–1401 (2008).
23. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
24. Raedt, T. D. *et al.* Conservation of hotspots for recombination in low-copy repeats associated with the *NF1* microdeletion. *Nature Genet.* **38**, 1419–1423 (2006).
25. Griffiths, R. C. & Tavaré, S. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math. Biosci.* **127**, 77–98 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to the participants who donated DNA samples, to D. Altshuler, J. Buard, K. Bryc, J. Kovacs, B. de Massy, G. McVean, B. Pasaniuc and S. Sankararaman for conversations and critiques, and to A. Auton for facilitating analysis of the 1000 Genomes Project data. Analysis was supported by the Wellcome Trust and NIH grants HL084107 and GM091332. CARE was supported by a contract from the National Heart, Lung and Blood Institute (HHSN268200960009C) to create a phenotype and genotype database for dissemination to the biomedical research community. Eight parent studies contributed phenotypic data and DNA samples through the Broad Institute (N01-HC-65226): the Atherosclerosis Risk in Communities study (ARIC), the Cleveland Family Study (CFS), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA) study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS) and the Sleep Heart Health Study (SHHS). Support for CARE also came from the individual research institutions, investigators, field staff and study participants. Individual funding information is available at <http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx>. All genome-wide genotyping of samples from the Children's Hospital of Pennsylvania (CHOP) was supported by an Institutional Development Award to the Center for Applied Genomics from the Children's Hospital of Philadelphia, a research award from the Landenberger Foundation and the Cotswold Foundation. We thank all study participants and the staff at the Center for Applied Genomics for performing the genotyping. The African American Breast Cancer Consortium (AABCC) was supported by a DoD Breast Cancer Research Program Era of Hope Scholar Award to C.A.H. and the Norris Foundation, and by grants to the component studies: MEC (CA63464, CA54281); CARE (HD33175); WCHS (CA100598, DAMD 170100334, Breast Cancer Research Foundation); SFBC (CA77305, DAMD 17966071); CBCS (CA58223, ES10126), PLCO (NCI Intramural Research Program); NHBS (CA100374); WFBC (R01-CA73629); and CPS-II (the American Cancer Society). The African American Prostate Cancer Consortium (AAPCC) was supported by grants CA63464, CA54281, CA1326792, CA148085 and HG004726, and by grants to the component studies: PLCO (NCI Intramural Research Program), LAAPC (Cancer Research Fund 99-00524V-10258), both MEC and LAAPC (PC35139, DP000807); MDA (CA68578, CA140388, ES007784, DAMD W81XWH0710645); GECAP (ES011126); CaP Genes (CA88164); IPCG (W81XWH0710122); DCPC (GM08016, DAMD W81XWH0710203, DAMD W81XWH0610066); and SCCS (CA092447, CA68485). The African American Lung Cancer Consortium (AALCC) was supported by grants CA060691, CA87895, PC35145 and CA22453, CA68578, CA140388, ES007784, ES06717, CA55769, CA127219, CA1116460S1, CA1116460, CA121197, CA141716, CA121197S2, CPRIT RP100443, CA148127, DAMD W81XWH0710645, University Cancer Foundation, Duncan Family Institute, Center for Community, Implementation and Dissemination Research Core, and by grants to the component studies: PLCO and the Maryland Studies (NCI Intramural Research Program), LAAPC (Cancer Research Fund 99-00524V-10258), and both MEC and LAAPC (PC35139, DP000807).

**Author Contributions** D.R. and S.R.M. conceived the study. A.G.H., A.T., N.P., Y.S., N.R., C.D.P., G.K.C., K.W., S.G.B., D.R. and S.R.M. performed analyses. N.R. performed the experimental work (genotyping of polymorphisms at *PRDM9*). A.G.H., N.P., J.N.H.,



B.E.H., H.A.T. Jr, A.L.P., H.H., S.J.C., C.A.H., J.G.W., D.R. and S.R.M. coordinated the study. A.G.H., D.R. and S.R.M. wrote the paper. N.R., C.D.P., G.K.C., K.W., S.G.B., S.R., J.N.H., B.E.H., H.A.T. Jr, H.H., S.J.C., C.A.H., J.G.W., D.R. and all the alphabetically listed authors contributed to sample collection and generation of SNP array data. All authors contributed to revision and review of the manuscript.

**Author Information** Crossover rate estimates for the AA map can be found at <http://www.well.ox.ac.uk/~anjali/AAmap/>. We also provide estimates of uncertainty for the map based on samples from the MCMC. Association testing results for each SNP are available from the authors on request. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.R. ([reich@genetics.med.harvard.edu](mailto:reich@genetics.med.harvard.edu)) or S.R.M. ([myers@stats.ox.ac.uk](mailto:myers@stats.ox.ac.uk)).

Anjali G. Hinch<sup>1</sup>, Arti Tandon<sup>2,3</sup>, Nick Patterson<sup>2</sup>, Yunli Song<sup>4</sup>, Nadin Rohland<sup>2,3</sup>, Cameron D. Palmer<sup>5,6</sup>, Gary K. Chen<sup>7</sup>, Kai Wang<sup>8,9</sup>, Sarah G. Buxbaum<sup>10</sup>, Ermeg L. Akylbekova<sup>10,11</sup>, Melinda C. Aldrich<sup>12,13</sup>, Christine B. Ambrosone<sup>14</sup>, Christopher Amos<sup>15</sup>, Elisa V. Bandera<sup>16</sup>, Sonja I. Berndt<sup>17</sup>, Leslie Bernstein<sup>18</sup>, William J. Blot<sup>13,19</sup>, Cathryn H. Bock<sup>20</sup>, Eric Boerwinkle<sup>21</sup>, Qiuyin Cai<sup>13</sup>, Neil Caporaso<sup>17</sup>, Graham Casey<sup>7</sup>, L. Adrienne Cupples<sup>22</sup>, Sandra L. Deming<sup>13</sup>, W. Ryan Diver<sup>23</sup>, Jasmin Divers<sup>24</sup>, Myriam Fornage<sup>25</sup>, Elizabeth M. Gillanders<sup>26</sup>, Joseph Glessner<sup>9</sup>, Curtis C. Harris<sup>27</sup>, Jennifer J. Hu<sup>28</sup>, Sue A. Ingles<sup>9</sup>, William Isaacs<sup>29</sup>, Esther M. John<sup>30</sup>, W. H. Linda Kao<sup>31</sup>, Brendan Keating<sup>9</sup>, Rick A. Kittles<sup>32</sup>, Laurence N. Kolonel<sup>33</sup>, Emma Larkin<sup>34</sup>, Loic Le Marchand<sup>33</sup>, Lorna H. McNeill<sup>35</sup>, Robert C. Millikan<sup>36</sup>, Adam Murphy<sup>37</sup>, Solomon Musani<sup>11</sup>, Christine Neslund-Dudas<sup>38</sup>, Sarah Nyante<sup>36</sup>, George J. Papanicolaou<sup>39</sup>, Michael F. Press<sup>7</sup>, Bruce M. Psaty<sup>40</sup>, Alex P. Reiner<sup>41</sup>, Stephen S. Rich<sup>42</sup>, Jorge L. Rodriguez-Gil<sup>28</sup>, Jerome I. Rotter<sup>43</sup>, Benjamin A. Rybicki<sup>38</sup>, Ann G. Schwartz<sup>20</sup>, Lisa B. Signorello<sup>13,19</sup>, Margaret Spitz<sup>15</sup>, Sara S. Strom<sup>44</sup>, Michael J. Thun<sup>23</sup>, Margaret A. Tucker<sup>17</sup>, Zhaoming Wang<sup>45</sup>, John K. Wiencke<sup>46</sup>, John S. Witte<sup>47</sup>, Margaret Wrensch<sup>46</sup>, Xifeng Wu<sup>15</sup>, Yuko Yamamura<sup>44</sup>, Krista A. Zanetti<sup>26,27</sup>, Wei Zheng<sup>13</sup>, Regina G. Ziegler<sup>17</sup>, Xiaofeng Zhu<sup>48</sup>, Susan Redline<sup>49</sup>, Joel N. Hirschhorn<sup>5,6,50</sup>, Brian E. Henderson<sup>7</sup>, Herman A. Taylor Jr<sup>11,51,52</sup>, Alkes L. Price<sup>53</sup>, Hakon Hakonarson<sup>9,54</sup>, Stephen J. Chanock<sup>17</sup>, Christopher A. Haiman<sup>7</sup>, James G. Wilson<sup>55</sup>, David Reich<sup>2,3\*</sup> & Simon R. Myers<sup>1,4\*</sup>

<sup>1</sup>Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford OX3 7BN, UK. <sup>2</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>3</sup>Department of Genetics, Harvard Medical School, New Research Building, 77 Ave. Louis Pasteur, Boston, Massachusetts 02115, USA.

<sup>4</sup>Department of Statistics, Oxford University, 1 South Parks Road, Oxford OX1 3TG, UK.

<sup>5</sup>Program in Medical and Population Genetics, Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>6</sup>Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital Boston, Massachusetts 02115, USA.

<sup>7</sup>Department of Preventive Medicine and Department of Pathology, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California 90033, USA. <sup>8</sup>Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, California 90089, USA. <sup>9</sup>Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. <sup>10</sup>Jackson Heart Study Coordinating Center, Jackson State University, 350 W. Woodrow Wilson Ave., Suite 701, Jackson, Mississippi 39213, USA. <sup>11</sup>Department of Medicine, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA.

<sup>12</sup>Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. <sup>13</sup>Division of Epidemiology in the Department of Medicine, Vanderbilt Epidemiology Center; and the Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37203, USA. <sup>14</sup>Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York 14263, USA.

<sup>15</sup>Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>16</sup>The Cancer Institute of New Jersey, New Brunswick, New Jersey 08903, USA. <sup>17</sup>Division of

Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. <sup>18</sup>Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, California 91010, USA. <sup>19</sup>International Epidemiology Institute, Rockville, Maryland 20850, USA. <sup>20</sup>Karmanos Cancer Institute and Department of Oncology, Wayne State University of Medicine, Detroit, Michigan 48201, USA. <sup>21</sup>Human Genetics Center and Division of Epidemiology, University of Texas at Houston, 1200 Herman Pressler St., Houston, Texas 77030, USA. <sup>22</sup>Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, Massachusetts 02118 and Framingham Heart Study, Framingham, Massachusetts 01702, USA.

<sup>23</sup>Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. <sup>24</sup>Department of Biostatistical Sciences, Wake Forest University School of Medicine WC-2326, Medical Center Blvd., Winston Salem, North Carolina 27157, USA. <sup>25</sup>Institute of Molecular Medicine and Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center at Houston, 1825 Pressler Street, Houston, Texas 77030, USA. <sup>26</sup>Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland 20892, USA. <sup>27</sup>Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA. <sup>28</sup>Sylvester Comprehensive Cancer Center and Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. <sup>29</sup>James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institutions, Baltimore, Maryland 21287, USA. <sup>30</sup>Cancer Prevention Institute of California, Fremont, California 94538; and Stanford University School of Medicine and Stanford Cancer Center, Stanford, California 94305, USA. <sup>31</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, Maryland 21205, USA.

<sup>32</sup>Department of Medicine, University of Illinois at Chicago, Chicago, Illinois 60607, USA. <sup>33</sup>Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. <sup>34</sup>Department of Medicine, Division of Allergy, Pulmonary and Critical Care, 6100 Medical Center East, Vanderbilt University Medical Center, Nashville, Tennessee 37232-8300, USA. <sup>35</sup>Department of Health Disparities Research, Division of OVP, Cancer Prevention and Population Sciences, and Center for Community Implementation and Dissemination Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>36</sup>Department of Epidemiology, Gillings School of Global Public Health, and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. <sup>37</sup>Department of Urology, Northwestern University, Chicago, Illinois 60611, USA. <sup>38</sup>Department of Public Health Sciences, Henry Ford Hospital, Detroit, Michigan 48202, USA. <sup>39</sup>Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute, 6701 Rockledge Drive, Bethesda, Maryland 20892, USA. <sup>40</sup>Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology & Health Services, University of Washington; Group Health Research Institute; Group Health Cooperative; 1730 Minor Ave., Seattle, Washington 98101, USA. <sup>41</sup>Department of Epidemiology, University of Washington, Box 357236 Seattle, Washington 98195, USA. <sup>42</sup>Center for Public Health Genomics, University of Virginia, West Complex Room 6111, Charlottesville, Virginia 22908, USA. <sup>43</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. <sup>44</sup>Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. <sup>45</sup>Core Genotype Facility, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland 20877, USA. <sup>46</sup>University of California San Francisco, San Francisco, California 94158, USA. <sup>47</sup>Institute for Human Genetics, Departments of Epidemiology and Biostatistics and Urology, University of California, San Francisco, San Francisco, California 94158, USA. <sup>48</sup>Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Wolstein Research Building, Cleveland, Ohio 44106, USA. <sup>49</sup>Brigham and Women's Hospital, Department of Medicine, Division of Sleep Medicine, 75 Francis Street, Boston, Massachusetts 02115, USA. <sup>50</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>51</sup>Jackson State University, 1400 Lynch Street, Jackson, Mississippi 39217, USA. <sup>52</sup>Tougaloo College, 500 West County Line Road, Tougaloo, Mississippi 39174, USA. <sup>53</sup>Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. <sup>54</sup>Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. <sup>55</sup>Department of Physiology and Biophysics, University of Mississippi Medical Center, 2500 N. State St., Jackson, Mississippi 39216, USA.

\*These authors contributed equally to this work.

## METHODS

**Samples used for building the AA map.** The 29,589 unrelated African-American samples derive from five sources. Informed consent was provided by all the individuals participating in the study, and was approved by all of the institutions responsible for sample collection.

The first source is the Candidate Gene Association Resource (CARE) study, a consortium of cohorts. We analysed CARE samples genotyped on the Affymetrix 6.0 array from the Atherosclerosis Risk in Communities study (ARIC), the Cleveland Family Study (CFS), the Coronary Artery Risk Development in Young Adults study (CARDIA), the Jackson Heart Study (JHS) and the Multi-Ethnic Study of Atherosclerosis (MESA). After removing individuals known to be related, and restricting to SNPs with good completeness in all cohorts, we had data from 6,209 individuals typed at 580,000 SNPs.

The second source consists of diverse studies carried out at the Children's Hospital of Philadelphia (CHOP), which has established a biobank for Philadelphia children to facilitate large genotype-phenotype association analysis. The cohort was recruited by CHOP clinicians, nursing and medical assistant staff within the CHOP Health Care Network, including primary care clinics and out-patient practices, from the hospital's patient base of over one million paediatric patients. All samples analysed here were genotyped on either the Illumina 610-Quad or Illumina HumanHap550 array. After removing individuals known to be related, identifying American Americans by multidimensional scaling on genotype data, and restricting to SNPs with a high level of completeness across samples, we had data from 7,503 samples typed at 491,572 SNPs.

The third source is the African American Breast Cancer Consortium (AABCC), consisting of the Multiethnic Cohort study (MEC), the Los Angeles component of the Women's Contraceptive and Reproductive Experiences study (CARE), the Women's Circle of Health Study (WCHS), the San Francisco Bay Area Breast Cancer study (SFBC), the Carolina Breast Cancer Study (CBCS), the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Cohort (PLCO), the Nashville Breast Health Study (NBHS) and the Wake Forest University Breast Cancer Study (WFBC), all genotyped on an Illumina 1M array. After data curation, including removal of samples with genetic evidence of being second-degree relatives or closer using the *smartrel* package of EIGENSOFT<sup>26</sup> ( $>0.2$  correlation of genotype state), we had data from 5,203 women (about half cases and half controls) typed at 894,717 SNPs.

The fourth source is the African American Prostate Cancer Consortium (AAPCC), consisting of the MEC, the Southern Community Cohort Study (SCCS), PLCO, the Cancer Prevention Study II Nutrition Cohort (CPS-II), the Prostate Cancer Case-Control Studies at MD Anderson (MDA), the Identifying Prostate Cancer Genes study (IPCG), the Los Angeles Study of Aggressive Prostate Cancer (LAAPC), the Prostate Cancer Genetics Study (CaP Genes), the Case-Control Study of Prostate Cancer among African Americans in Washington DC (DCPC), the Gene-Environment Interaction in Prostate Cancer Study (GECAP) and the Cancer Prevention Study II (CPS-II), all typed on an Illumina 1M array. After the same data curation as the breast cancer study, we had data from 6,540 men (about half cases and half controls) typed at 896,036 SNPs.

The fifth source is individuals from the African American Lung Cancer Consortium (AALCC), including cases and controls from the MEC, the SCCS, PLCO, the MD Anderson (MDA) African American Lung Cancer Study, the NCI-Maryland Lung Cancer Case-Control Study, the University of California at San Francisco African American Lung Cancer Study and the Wayne State African American Lung Cancer Study, all genotyped on the Illumina 1M array. After data curation, we had data from 4,134 individuals typed at 906,687 SNPs.

**Samples used for building the pedigree map.** The pedigree map was built using data from 135 African-American nuclear families from CARE and 87 African-American families from CHOP for which genotyping data were available from at least two full siblings and at least one parent. The CARE studies that contributed samples were JHS (70 families, including 58 samples that we newly genotyped on the Affymetrix 6.0 array to increase the number of crossovers we could analyse) and CFS (65 families). For the families with a missing parent, we developed a Hidden Markov Model (HMM) approach to jointly estimate the genotype of the missing parent as well as to infer the position of crossover events in the offspring. The observed variables in the HMM were the genotypes of the available family members and the states of the HMM were the genotypes of the parents and the identity by descent (IBD) status of the children. A change in IBD status in an offspring is interpreted as a crossover event. Supplementary Note 2 provides details of the HMM used to infer positions of these pedigree crossover events.

**Local ancestry inference and identification of crossover events.** We merged the data for each cohort with phased YRI and CEU data from the HapMap3 data set<sup>27</sup>. We filtered SNPs that had a frequency inconsistent with an 80–20% linear combination of YRI and CEU frequencies ( $t$  statistic with an absolute value of greater

than 3), potentially reflecting genotyping error in either the HapMap3 or the cohort data.

We ran HAPMIX on these data using a prior hypothesis of 20% European ancestry and 6 generations since mixture for each individual<sup>15</sup>. HAPMIX requires users to input a recombination map as a prior distribution, and we assumed that rates were constant across each chromosome arm with a total rate across each arm determined by the Rutgers genetic map<sup>6</sup> (Supplementary Note 1).

Filtering of crossover events had three stages. First, we removed crossover events where the probability of occurrence was estimated to be less than 95% by HAPMIX. Second, we removed candidate crossover events that were non-monotonic, that is, where the probability of an overlapping crossover event with an ancestry switch in a different direction was  $\geq 1\%$  within any inter-SNP interval. Third, we removed crossover events where either of the two flanking ancestry blocks was smaller than 2 cM in size as measured with respect to a published map based on linkage disequilibrium<sup>3,18</sup> (Supplementary Note 1). For comparisons to the deCODE map and linkage-disequilibrium-based maps, we also removed segments of the genome within 5 Mb of the telomeres (to be consistent with the comparisons presented in the deCODE study where the same restriction was applied<sup>4</sup>).

**Construction of the AA map.** All 22 autosomes and chromosome X were split into approximately 1.3 million inter-SNP intervals based on the union of SNPs analysed across all five sample sets. Our goal was to estimate a crossover rate for each of these intervals. We modelled crossover rates such that the rate for each SNP interval is independent of every other SNP interval, motivated by a hotspot model. We used a gamma prior on rates with the mean estimated from the filtered HAPMIX output (Supplementary Note 1). We used a Gibbs sampler to sample rates in every SNP interval and to determine the location of a crossover event within the 95% range estimated by the HAPMIX output. In each round of the Gibbs sampler, we used the set of sampled rates in the previous round to construct a probability mass function for the SNP interval in which each crossover occurred, using an approach described in Supplementary Note 1 to approximate the probability mass function that HAPMIX would have produced conditional on the previous set of sampled rates. After sampling the location of the crossover events, we counted how many crossovers occurred in every SNP interval. We used these counts to construct a posterior distribution for the crossover rate in each SNP interval, taking advantage of the conjugacy of a Poisson likelihood and a gamma prior. We then sampled a crossover rate for each SNP interval from its respective gamma posterior distribution.

**Candidate African-enriched hotspots.** To identify candidate African-enriched hotspots, we used two pairs of maps: the previously available YRI map and CEU map, and the AE map and the S map. We combined information from both map pairs to enrich for regions with genuine differences between the West African and European populations. Specifically, we identified candidate hotspots as 2-kb intervals representing a peak in the AE map rate, where the estimated rate in the AE map was  $>2 \text{ cM Mb}^{-1}$  and at least double that in the S map, and in addition the YRI map rate was  $>2 \text{ cM Mb}^{-1}$  and at least double the CEU map rate. We took the resulting candidate hotspot set and defined hotspot boundaries by identifying the region flanking the 2 kb rate peak that had rates at least 50% of the peak value in the AE map. Regions larger than 5 kb were discarded. We similarly constructed a set of 'shared' hotspots but modified the initial criteria given the lack of obvious hotspots present only in people of European ancestry. Specifically, we identified 2 kb S map rate peak locations where both the S and CEU estimated rates were  $>2 \text{ cM Mb}^{-1}$ , while the AE and YRI map rates were below those in these respective European populations. We then narrowed the regions and filtered using the same procedure we had developed for the candidate African-enriched hotspots.

**Association testing.** MaCH<sup>28</sup> was used to impute up to 3,058,149 SNP genotypes from HapMap2 (ref. 18) into all African Americans we analysed, using the unrelated YRI and CEU samples as combined reference panels. We tested for association at all SNPs with minor allele frequency  $> 1\%$ . To restrict our analysis to individuals in whom the phenotype was measured accurately, we performed the association analysis with the AE and hotspot usage phenotypes only in individuals with at least 35 inferred crossovers. Association testing was carried out using linear regression, after controlling for gender, genome-wide European ancestry proportion (inferred by HAPMIX) and study (Supplementary Note 4). We observe slight inflation of the association statistics genome-wide compared with the expectation (the Genomic Control inflation factor<sup>29</sup> is 1.046 for the AE phenotype and 1.038 for the hotspot usage phenotype), which we propose may reflect cryptic relatedness among samples (Supplementary Note 4). We report  $P$  values after correction using Genomic Control<sup>29</sup>.

**Construction of PRDM9 tree.** To examine the history of the PRDM9 ZF array and to place SNPs showing association with AE map usage within the framework of this history, we identified 19 SNPs from HapMap2 (ref. 18) that surrounded the

ZF array and that form a maximal block of SNPs where there is almost no evidence of recombination:  $|D'| = 1$  for all pairs of SNPs in the data after removing 2 of 120 YRI and 1 of 120 CEU haplotypes (the chimpanzee genome was used to define the ancestral alleles). A unique 'gene tree' was then built, and we used *genetree*<sup>25</sup>, which assumes a coalescent prior on genealogies, to approximately infer ages for these mutations conditional on the data (a caveat is that the tree building does not account for the HapMap SNP ascertainment scheme). Because *genetree* assumes a randomly mating population, and the YRI represent almost all the HapMap haplotype diversity in this region, we ran the software (2,000,000 importance samples, otherwise default parameters) on the YRI data only and used this to construct Fig. 2c. Each node of the tree corresponds to a unique haplotype at these 19 SNPs, whose frequency in both CEU and YRI is shown at the base of the figure.

**Motif searching.** We tested all candidate motifs of 5 to 9 base pairs for enrichment in our African-enriched hotspot set relative to our shared hotspot set. We counted occurrences of all tested motifs in repeat and non-repeat backgrounds separately, and computed a separate *P* value for each genomic background with a chi-squared test, based on a contingency table that compares the counts of a particular motif to the counts of all motifs of that size. We converted each *P* value to a *Z* score, added the scores on each background, and then obtained a corresponding combined *P* value. Motifs were considered statistically significant only if they passed four stringent criteria: (1) they were statistically significant after Bonferroni correction for the number of motifs tested; (2) they were overrepresented in the African-enriched set; (3) they were statistically significant on both the repeat and non-repeat backgrounds ( $P < 0.01$ ) independently; and (4) they were statistically significant when the joint *P* value was calculated only by comparing the frequency of the motif to other motifs of identical G/C content (to eliminate false positives due to any difference in G/C content between the hotspot sets). This testing revealed a unique significant motif, the 9-nucleotide oligomer CCCCAGTGA. We explored whether flanking DNA around exact matches to this motif also had a role by testing whether bases at a given site relative to the motif were associated with the difference in rates between African- and European-ancestry populations (Kruskal–Wallis test). Rates were evaluated in the 2 kb surrounding each motif occurrence. We separately evaluated flanking sequence using both the difference between YRI/CEU map rates, and the difference between the AE/S map rates, leading to the identification of the 17-bp

consensus African-enriched motif (Supplementary Note 6 has full details). To identify close matches to this 17-bp motif among all matches to the 9-bp motif in the genome, for every occurrence of the 9-bp motif, we scored the flanking sequence bases proportionately to the relative increase in average crossover rate difference associated with each base, then multiplied across bases in the 17-mer region to provide an overall score. We ranked occurrences according to this score, and plotted rates around the top 500 (Fig. 3b). We verified these findings by measuring average crossover differences for each base using only odd chromosomes and used these to score motif occurrences on the (non-overlapping) set of even chromosomes, and vice versa (Supplementary Fig. 8).

**PRDM9 ZF length typing and genotyping of rs6889665.** To determine the number of ZF motifs of *PRDM9* in a subset of the samples used to build the map, published primer pairs<sup>4</sup> were used to amplify this region (forward: 5'-GGCCAGAAAGTGAATCCAGG-3', reverse: 5'-GGGGAATATAAGGGGTCAGC-3'). Product lengths ranged between 7 and 20 repeats (801–1,893 bp). Four of the 166 African-American samples did not show an amplification product, presumably because of insufficient DNA quality. We also genotyped 90 YRI and 90 CEU HapMap samples.

The SNP rs6889665 was genotyped in the same samples using an allelic discrimination assay (forward primer: 5'-aaacttggaacatccatagggt-3', reverse primer: 5'-cgaaaggagaaaagcataatcc-3', Locked Nucleic Acid (LNA) probe 'C': 5'-/6-FAM/aGGGatAaatgaag/BHQ/-3', LNA-probe 'T': 5'-/HEX/AGAGatAaatGaagg/BHQ/-3'; LNA bases are given in capital letters). Reporter dyes: 6-FAM, 6-carboxyfluorescein; HEX, hexachlorofluorescein. Quencher: BHQ, Black Hole Quencher 1. Only one out of the 166 African-American samples failed in this assay. The same YRI and CEU samples as above were also genotyped.

26. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
27. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
28. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
29. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).



# Functional genomics reveal that the serine synthesis pathway is essential in breast cancer

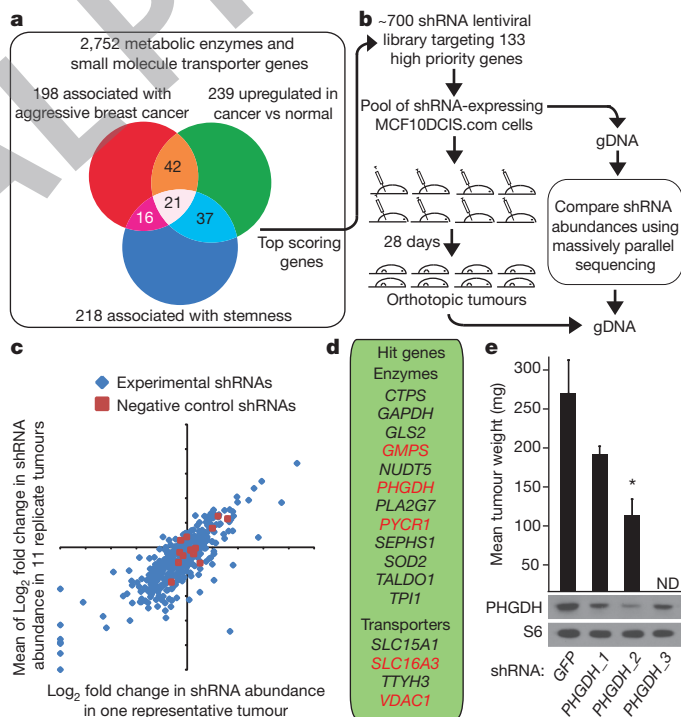
Richard Possemato<sup>1,2,3,4</sup>, Kevin M. Marks<sup>5</sup>, Yoav D. Shaul<sup>1,2,3,4</sup>, Michael E. Pacold<sup>1,2,3,4,6</sup>, Dohoon Kim<sup>1,2,3,4</sup>, Kıvanç Birsoy<sup>1,2,3,4</sup>, Shalini Sethumadhavan<sup>5</sup>, Hin-Koon Woo<sup>5</sup>, Hyun G. Jang<sup>5</sup>, Abhishek K. Jha<sup>5</sup>, Walter W. Chen<sup>1,2,3,4</sup>, Francesca G. Barrett<sup>1</sup>, Nicolas Stransky<sup>3</sup>, Zhi-Yang Tsun<sup>1,2,3,4</sup>, Glenn S. Cowley<sup>3</sup>, Jordi Barretina<sup>3,7</sup>, Nada Y. Kalaany<sup>1,2,3,4</sup>, Peggy P. Hsu<sup>1,2,3,4</sup>, Kathleen Ottina<sup>1,2,3,4</sup>, Albert M. Chan<sup>1,2,3,4</sup>, Bingbing Yuan<sup>1</sup>, Levi A. Garraway<sup>3,7</sup>, David E. Root<sup>3</sup>, Mari Mino-Kenudson<sup>8</sup>, Elena F. Brachtel<sup>8</sup>, Edward M. Driggers<sup>5</sup> & David M. Sabatini<sup>1,2,3,4</sup>

Cancer cells adapt their metabolic processes to drive macromolecular biosynthesis for rapid cell growth and proliferation<sup>1,2</sup>. RNA interference (RNAi)-based loss-of-function screening has proven powerful for the identification of new and interesting cancer targets, and recent studies have used this technology *in vivo* to identify novel tumour suppressor genes<sup>3</sup>. Here we developed a method for identifying novel cancer targets via negative-selection RNAi screening using a human breast cancer xenograft model at an orthotopic site in the mouse. Using this method, we screened a set of metabolic genes associated with aggressive breast cancer and stemness to identify those required for *in vivo* tumorigenesis. Among the genes identified, phosphoglycerate dehydrogenase (PHGDH) is in a genomic region of recurrent copy number gain in breast cancer and PHGDH protein levels are elevated in 70% of oestrogen receptor (ER)-negative breast cancers. PHGDH catalyses the first step in the serine biosynthesis pathway, and breast cancer cells with high PHGDH expression have increased serine synthesis flux. Suppression of PHGDH in cell lines with elevated PHGDH expression, but not in those without, causes a strong decrease in cell proliferation and a reduction in serine synthesis. We find that PHGDH suppression does not affect intracellular serine levels, but causes a drop in the levels of  $\alpha$ -ketoglutarate, another output of the pathway and a tricarboxylic acid (TCA) cycle intermediate. In cells with high PHGDH expression, the serine synthesis pathway contributes approximately 50% of the total anaplerotic flux of glutamine into the TCA cycle. These results reveal that certain breast cancers are dependent upon increased serine pathway flux caused by PHGDH overexpression and demonstrate the utility of *in vivo* negative-selection RNAi screens for finding potential anticancer targets.

As a starting point for identifying metabolic genes required for tumorigenesis, we cross-referenced maps of metabolic pathways with the KEGG database to compile a comprehensive list of 2,752 genes encoding all known human metabolic enzymes and transporters (Supplementary Table 1). Public oncogenomic data were analysed to score genes based on three properties: (1) higher expression in tumours versus normal tissues; (2) high expression in aggressive breast cancer; or (3) association with the stem-cell state (Fig. 1a). Genes scoring in two of these three categories as well as those at the top of each category were selected to define a high-priority set of 133 metabolic enzyme and transporter genes (Supplementary Table 2). We assembled lentiviral short hairpin RNA (shRNA) vectors targeting these genes (median 5 shRNAs per gene) and used them to generate two libraries of shRNA-expressing lentiviruses, one containing 235 distinct shRNAs (targeting

transporters and control genes) and the other 516 distinct shRNAs (targeting metabolic enzymes and control genes)<sup>4</sup>.

To identify genes that may be essential for tumorigenesis, the libraries were screened for shRNAs that became depleted during breast tumour formation in mice. Human MCF10DCIS.COM cells<sup>5</sup> were chosen for the screens because, of several breast cancer lines examined, these were capable of forming tumours upon injection of the fewest number of cells. One and a half million MCF10DCIS.COM cells were infected with each library so that each cell carried one viral integrant, and



**Figure 1 | Outline of *in vivo* pooled screening strategy identifying PHGDH as essential for tumorigenesis.** **a**, Venn Diagram outlining meta-analysis. **b**, Outline of experimental design. gDNA, genomic DNA. **c**, Log<sub>2</sub> fold change in shRNAs abundance of experimental (blue) or neutral shRNAs (red) for a single tumour (x-axis) compared to an average of eleven tumours (y-axis). **d**, Genes scoring *in vivo*. **e**, Average weight of tumours from MCF10DCIS.COM cells expressing shRNAs targeting PHGDH (PHGDH\_1, PHGDH\_2 and PHGDH\_3) or control (GFP) and protein expression of PHGDH or RPS6 (S6). Error bars are s.e.m. ( $n = 4$ ). \* $P$  value  $< 0.05$ . ND, not done.

<sup>1</sup>Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>2</sup>Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Broad Institute of Harvard and MIT, Seven Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>4</sup>The David H. Koch Institute for Integrative Cancer Research at MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Ajios Pharmaceuticals, 38 Sidney Street, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Harvard Radiation Oncology Program, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02114, USA. <sup>7</sup>Department of Medical Oncology and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute and Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. <sup>8</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA.

~500–1,000 cells per shRNA (100,000–1,000,000 cells total) were injected into mouse mammary fat pads at two sites per animal (Supplementary Discussion). Twenty-eight days later orthotopic tumours were harvested and massively parallel DNA sequencing was used to determine the abundance of each shRNA in genomic DNA from tumours and initially injected cells (Fig. 1b). shRNA abundances correlated well between replicate tumours (Fig. 1c) and 5 or 12 tumours per library were analysed to identify shRNAs that became significantly depleted during tumour formation. Sixteen genes were designated hits in the screen, with at least 75% of the shRNAs targeting these genes scoring (Fig. 1d and Supplementary Table 3).

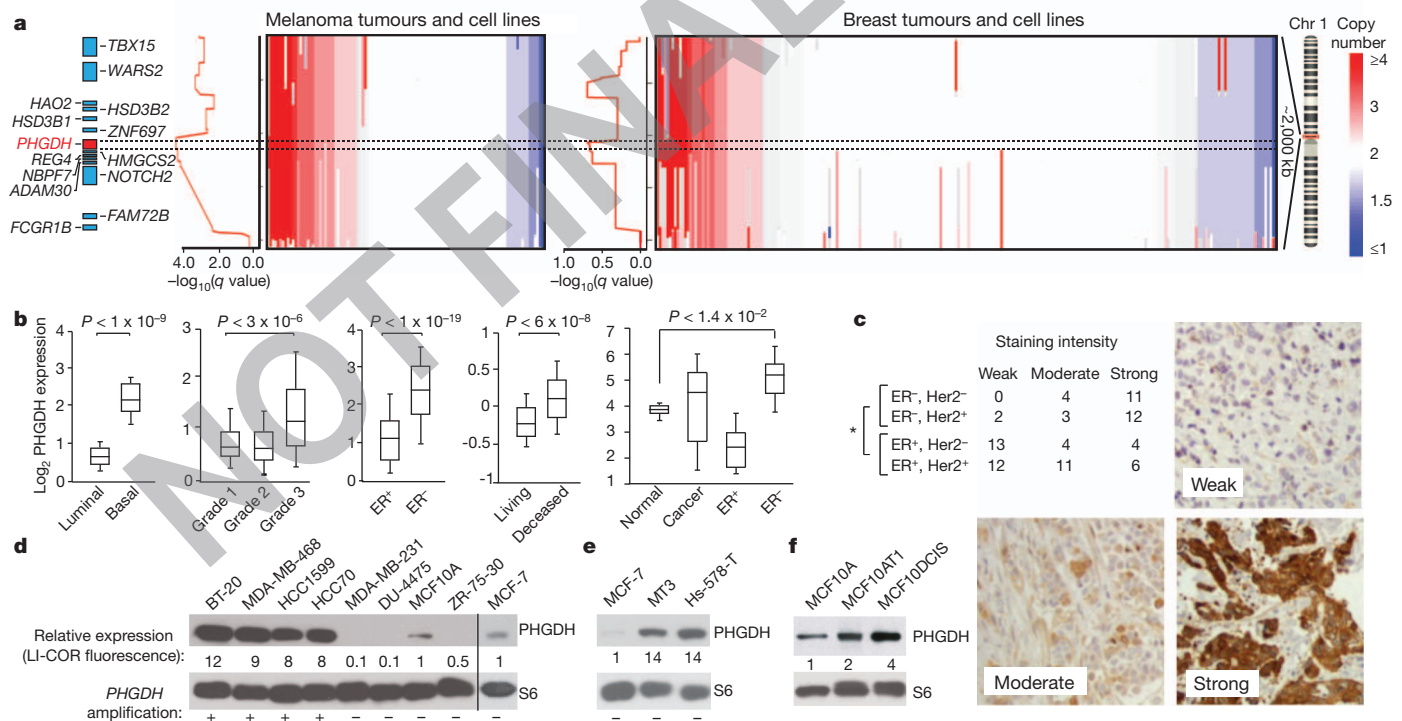
Several genes previously shown to have important roles in cancer emerged as hits, including the mitochondrial ATP transporter *VDAC1*; the lactic acid transporter *SLC16A3*; and the nucleotide synthesis genes *GMPS* and *CTPS*. The hit list also includes genes involved in the control of oxidative stress (*SOD2*, *GLS2*, *SEPHS1*), the pentose phosphate pathway (*TALDO1*), glycolysis (*GAPDH*, *TPI1*), and in the proline (*PYCR1*) and serine (*PHGDH*) biosynthetic pathways. An analogous pooled screen carried out in MCF10DCIS.com cells grown in culture rather than in tumour xenografts revealed that of 20 genes that scored in the *in vitro* screen, 10 also scored in the *in vivo* screen (Supplementary Fig. 2a, Supplementary Table 3 and Supplementary Discussion). Interestingly, *AK2*, which encodes an adenylate kinase that generates ADP from ATP and AMP, was required for *in vitro* but not *in vivo* growth (Supplementary Fig. 2b).

For five hit genes (*PHGDH*, *GMPS*, *SLC16A3*, *PYCR1* and *VDAC1*), two scoring shRNAs were tested for their effects on tumour formation. Each of these shRNAs suppressed expression of their targets in MCF10DCIS.com cells and reduced tumour-forming capacity.

(Fig. 1e and Supplementary Fig. 2c). For reasons discussed later, *PHGDH* was of particular interest. The three shRNAs that scored in the *in vivo* screen also decreased *PHGDH* protein expression, and two shRNAs of differing knockdown efficacies inhibited tumour growth, consistent with their capacity to suppress *PHGDH* expression (Fig. 1e). Moreover, tumours derived from cells that in culture had confirmed reductions in *PHGDH* levels had, in immunohistochemical (Supplementary Fig. 3a) and immunoblotting assays (Supplementary Fig. 3b), *PHGDH* staining or levels similar to control tumours, suggesting that tumorigenesis selected for cells that lost shRNA-mediated *PHGDH* suppression.

To prioritize genes for follow-up studies we consulted a recently available analysis of copy number alterations across cancer genomes<sup>6</sup>. Indeed, *PHGDH* exists in a region of chromosome 1p commonly amplified in breast cancer and melanoma (Fig. 2a), as well as in several other cancer types (not shown). In total, 18% of patient-derived breast cancer cell lines and 6% of primary tumours have amplifications in *PHGDH*. In the data sets examined, none of the other hit genes are in genomic regions of focal and recurrent copy number gain.

Our meta-analysis for genes associated with aggressive breast cancer is corroborated by a previous study that found elevated *PHGDH* messenger RNA levels in breast cancers that are ER negative, of the basal type, and associated with poor 5-year survival<sup>7</sup>. We confirmed these associations in distinct gene expression data sets (Fig. 2b) and additionally found that *PHGDH* is elevated in ER-negative breast cancer relative to normal breast tissue (Fig. 2b). Of all the genes identified as hits in our screen, *PHGDH* has the most significantly elevated expression in ER-negative breast cancer (Supplementary Fig. 4). Moreover, by analysing 82 human breast tumour samples with an



**Figure 2 | Genomic amplifications of *PHGDH* in cancer and association of *PHGDH* expression with aggressive breast cancer markers.** **a**, *PHGDH* vicinity copy number data for melanoma (left,  $n = 111$ ) and breast cancer (right,  $n = 243$ ) samples. Coloured bar indicates degree of copy number loss (blue) or gain (red). Samples sorted by copy number at *PHGDH* locus (dotted lines). Graphs at left of copy number data show amplification significance ( $-\log_{10}(q \text{ value})$ ), ~0.60 is the significance threshold for amplification. **b**, Representative *PHGDH* gene expression data for indicated breast cancer groups. Whiskers indicate 91st and 9th percentile. **c**, Table reports numbers of

human breast cancer samples with 'weak', 'moderate', or 'strong' *PHGDH* staining from breast cancer subgroups indicated. Representative staining intensities shown in images. Magnification,  $\times 20$ . \* $P < 0.0001$  comparing ER<sup>-</sup> versus ER<sup>+</sup> classes (Fisher's exact test). **d–f**, *PHGDH* protein levels are shown for *PHGDH* amplified versus non-amplified (annotated with + or -) (**d**), *PHGDH* non-amplified, over-expressing (**e**), and MCF10A-derived cell lines (**f**). Values below *PHGDH* immunoblots are normalized immunofluorescent quantification (LI-COR) of *PHGDH* levels relative to actin control and MCF-10A and MCF7.

immunohistochemical assay for PHGDH, we found that PHGDH protein levels correlate significantly with ER-negative status (Fig. 2c). In total, compared to ER-positive breast tumours, ~68% and ~70% of ER-negative breast tumours have elevations of PHGDH at the mRNA and protein levels, respectively (Fig. 2b, c and Supplementary Methods). ER-negative breast cancer comprises approximately 20–25% of all breast cancer cases, but as many as 50% of all breast cancer deaths within 5 years of diagnosis<sup>8</sup>, underscoring the importance of identifying additional drug targets for this class of breast cancer.

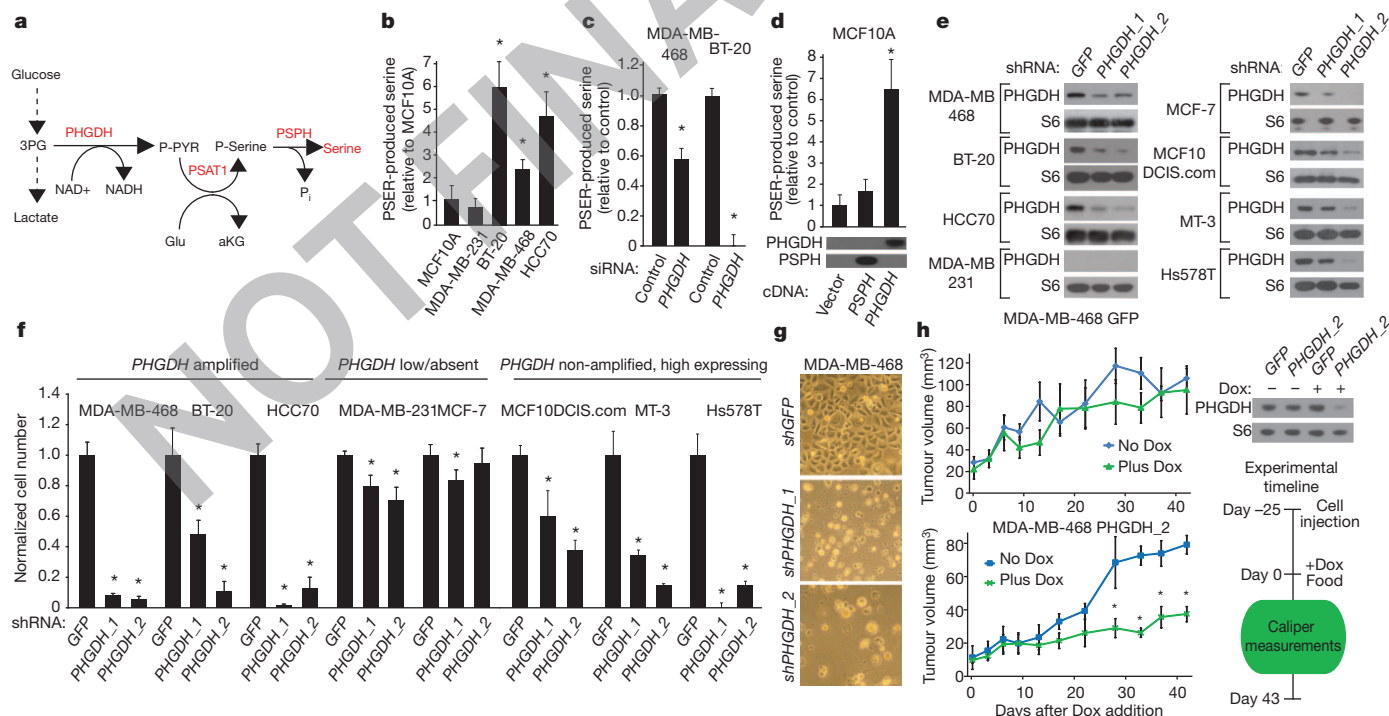
Across a set of breast cancer lines, four lines with *PHGDH* amplifications had 8–12-fold higher PHGDH protein expression compared to non-transformed MCF10A and ER-positive MCF7-cell lines, which do not have *PHGDH* amplifications (Fig. 2d). Mechanisms other than gene copy number increases must also exist for boosting PHGDH expression because PHGDH protein levels were also elevated in two ER-negative cell lines (MT3, Hs578T) lacking the *PHGDH* amplification (Fig. 2e). This is consistent with the finding that PHGDH expression is upregulated at the mRNA and protein level in a higher fraction of ER-negative breast cancers than the fraction exhibiting amplification at the DNA level. Interestingly, PHGDH is also expressed fourfold more in the MCF10DCIS.COM cells used in the *in vivo* screen than in two parental lines (MCF-10A and MCF10AT) that exhibit no or lower tumorigenicity<sup>9</sup> (Fig. 2f).

*PHGDH* encodes 3-phosphoglycerate dehydrogenase, the first enzyme branching from glycolysis in the three-step serine biosynthetic pathway<sup>10</sup> (Fig. 3a). PHGDH uses NAD as a cofactor to oxidize the glycolytic intermediate 3-phosphoglycerate into phosphohydroxypyruvate<sup>11,12</sup>, which subsequent enzymes in the pathway convert into serine via transamination (PSAT1) and phosphate ester hydrolysis (PSPH) reactions<sup>10</sup> (Fig. 3a). Serine is essential for synthesis of proteins and other biomolecules needed for cell proliferation, including nucleotides, phosphatidyl-serine and sphingosine

(Supplementary Fig. 1). Classic studies show elevated serine biosynthetic activity, as determined by enzyme assays, in rat tumour lysates<sup>10,13</sup>, and suggest that PSPH is the rate-limiting enzyme of this pathway in the liver<sup>14</sup>. Interestingly, we find that numerous genes that are expected to promote serine biosynthesis or are involved in the subsequent metabolism of serine for biosynthesis are elevated in ER-negative breast cancer (Supplementary Fig. 5), demonstrating that PHGDH elevation occurs in the context of upregulation of a broader pathway.

To understand the metabolic consequences of increased PHGDH expression we used metabolite profiling and serine synthesis pathway flux analysis to examine breast cancer cells with and without *PHGDH* amplifications. We found that cells with *PHGDH* amplifications (BT-20, MDA-MB-468 and HCC70), had increased flux through the serine synthesis pathway compared to those without *PHGDH* amplifications (MDA-MB-231, MCF7 and MCF10A) (Fig. 3b and Supplementary Fig. 6a). Cells with elevated PHGDH and high pathway flux were capable of robust proliferation in medium lacking serine, whereas in cells with low levels of PHGDH, the deprivation of serine caused a significant blunting or even cessation of proliferation (Supplementary Fig. 6b).

PHGDH is required for the increased serine pathway flux of cells with elevated PHGDH because RNAi-mediated PHGDH suppression significantly reduced flux in MDA-MB-468 and BT-20 cells (Fig. 3c). Conversely, in MCF-10A human mammary cells engineered to over-express PHGDH, serine pathway flux increased to levels similar to those in MDA-MB-468, BT-20 and HCC70 cells (Fig. 3d). Furthermore, MCF-10A cells overexpressing PHGDH had increased proliferation in the absence of serine, indicating that PHGDH overexpression is sufficient to drive flux through the pathway (Supplementary Fig. 6c). Interestingly, overexpression of PSPH, considered the rate-limiting serine biosynthetic enzyme in the liver, did not increase pathway flux in MCF-10A cells (Fig. 3d). The observation that PSPH is rate limiting



**Figure 3 | Cell lines with elevated PHGDH expression have increased serine biosynthetic pathway activity and are sensitive to PHGDH suppression.**

**a**, Serine biosynthesis pathway. **b–d**, Serine production by serine biosynthesis pathway in indicated breast cell lines (**b**), after PHGDH suppression by siRNA (**c**), and MCF-10A cells expressing *PHGDH* or *PSPH* cDNAs with associated immunoblots (**d**). **e**, Immunoblots of indicated proteins for indicated cell lines expressing control shRNA (*GFP*) or shRNAs against PHGDH (*PHGDH\_1* and *PHGDH\_2*). **f**, Relative proliferation of cells transduced with shRNA constructs

after seven days. **g**, Images showing cellular morphology (magnification, ×20) of MDA-MB-468 at day seven of **f**. **h**, Tumour growth of MDA-MB-468 cells expressing doxycycline-inducible control shRNA (*GFP*) or shRNA against *PHGDH* (*shPHGDH\_2*) in mice fed doxycycline (Dox, 2 mg kg<sup>-1</sup>, green lines, *n* = 5) or normal (blue lines, *n* = 4) diet after initial tumour palpation (day 0). Immunoblots of PHGDH or RPS6 (S6) shown for cells *in vitro*. \**P* < 0.05 relative to control. Error bars for metabolite measurements (*n* = 4) and tumour size indicate s.e.m., and for cell number indicate s.d. (*n* = 3).



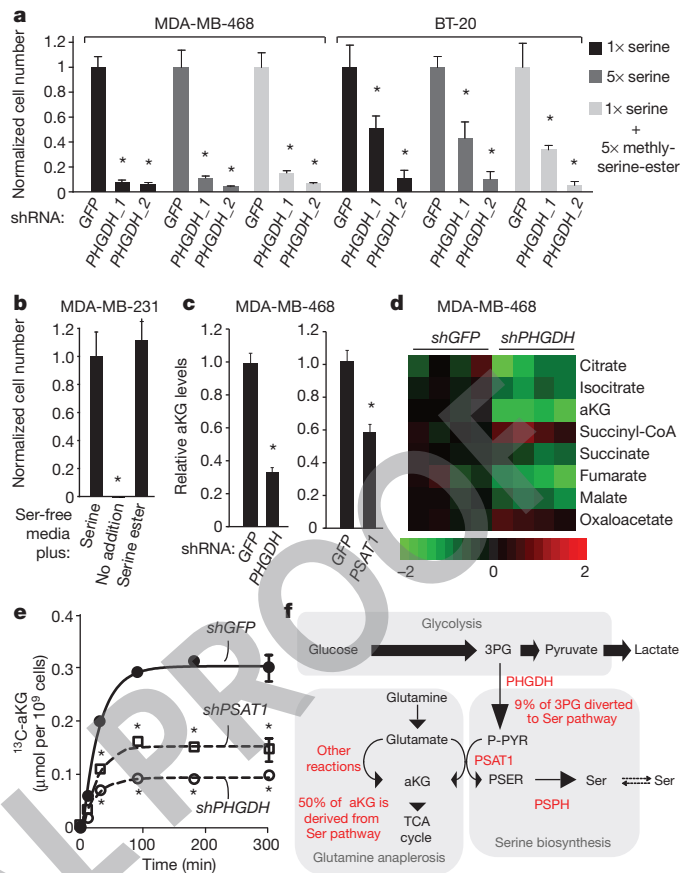
in the liver whereas PHGDH is rate limiting in MCF10A cells can be reconciled by the observation that serine levels in the liver (2 mM) are well above the concentration at which PSPH is feedback-inhibited by serine (500  $\mu$ M), but low in cell lines in culture ( $\sim$ 100  $\mu$ M), a concentration at which PSPH should be active<sup>14</sup>. These data demonstrate that PHGDH is a key enzyme controlling flux through the serine biosynthetic pathway in cancer cells.

We next asked if cells with an increase in PHGDH expression require it for cell proliferation and survival. In cell lines with elevated PHGDH expression (BT-20, MDA-MB-468, HCC70, Hs578T and MT3), but not without (MDA-MB-231 and MCF-7), RNAi-mediated suppression of PHGDH caused a marked decrease in cell number (Fig. 3e, f and Supplementary Fig. 6d) and cell death (Fig. 3g and Supplementary Fig. 6e) in the absence of apoptotic markers (Supplementary Fig. 6f). This sensitivity to PHGDH suppression was observed both in cells with *PHGDH* amplifications (BT-20, MDA-MB-468 and HCC70) and in those with high PHGDH expression but lacking *PHGDH* amplification (MT3 and Hs578T). Consistent with flux through the serine synthesis pathway being important in cells with high PHGDH expression, suppression of the other two enzymes in the pathway (PSAT1 and PSPH) inhibited the proliferation of MDA-MB-468 and BT-20, but not MCF7, cells (Supplementary Fig. 6g). Moreover, inhibition of PSPH inhibited tumour formation by MCF10DCIS.com cells (Supplementary Fig. 6h). Therefore, elevated PHGDH expression defines a set of breast cancer cell lines with increased serine pathway flux that are dependent upon PHGDH, PSAT1 and PSPH for proliferation. This finding suggests that many ER-negative breast cancers that express PHGDH at high levels ( $\sim$ 70% of all ER-negative disease in our data set; Fig. 2c) may be sensitive to inhibitors of the serine synthesis pathway.

To investigate whether PHGDH suppression can affect the growth of established tumours, we generated an inducible shRNA<sup>15</sup> that, upon doxycycline treatment, reduced PHGDH protein levels in MDA-MB-468 cells (Fig. 3h). MDA-MB-468 cells transduced with this shRNA were allowed to form murine mammary fat pad tumours for 25 days before introduction of doxycycline in a subset of mice (Fig. 3h). Compared to control mice, those given doxycycline exhibited substantially reduced tumour growth, whereas tumours made from cells transduced with a control inducible shRNA grew equally well in the presence or absence of doxycycline (Fig. 3h). These results indicate that PHGDH suppression can adversely affect growth in existing tumours (Supplementary Discussion).

Serine is a central metabolite for biosynthetic reactions, and we find that overexpression of PHGDH contributes significantly to biosynthetic flux to serine. However, PHGDH suppression inhibited proliferation even in cells growing in media containing normal levels of extracellular serine (Fig. 3f), and supplementation with additional serine or a cell-permeable methyl-serine-ester did not blunt the effects of the PHGDH suppression (Fig. 4a, b). Intracellular and extracellular serine are in equilibrium (Supplementary Fig. 7a), and import of extracellular serine was not defective in the cell lines studied (Supplementary Fig. 7b). These findings suggest that serine production may not be the only important role of PHGDH in cell lines with high PHGDH expression. We considered three hypotheses to explain our observations: (1) serine produced via the PHGDH pathway is used in a different manner than exogenous serine; (2) suppression of PHGDH adversely affects glycolysis; or (3) the PHGDH, PSAT1 and PSPH reactions produce metabolites besides serine that are also critical for cell proliferation. The first hypothesis was deemed unlikely because serine synthesized intracellularly is in equilibrium with extracellular serine (Supplementary Fig. 7a). The second hypothesis was also unlikely because PHGDH suppression did not affect glucose uptake or lactate production (Supplementary Fig. 7c).

To pursue the third hypothesis, we considered which additional metabolites the serine synthesis pathway might produce in significant levels in cells with high PHGDH expression. The serine pathway



**Figure 4 | Suppression of PHGDH results in a deficiency in anaplerosis of glutamine to aKG.** **a**, Relative proliferation of cell lines indicated expressing control shRNA (GFP) or shRNAs against *PHGDH* (*PHGDH\_1* and *PHGDH\_2*) after seven days of growth under conditions indicated. **b**, Relative proliferation of MDA-MB-231 cells under conditions indicated. **c**, Intracellular aKG four days after treatment with shRNA against *PHGDH* or *PSAT1*; cell number normalized relative to control shRNA (GFP). **d**, TCA cycle intermediate levels four days after treatment with shRNA against *PHGDH* or GFP ( $n = 4$ ). Colour bar shows Log<sub>2</sub> scale. **e**, aKG isotopic labelling at indicated time points after treatment with isotopically labelled glutamine four days after treatment with shRNA against *PHGDH*, *PSAT1* or GFP. **f**, Model of relative metabolite fluxes for indicated pathways. \* $P < 0.05$  relative to control. Error bars indicate s.e.m. ( $n = 4$ ).

produces equimolar amounts of serine and  $\alpha$ -ketoglutarate (aKG; Supplementary Fig. 1). Proliferating cells use intermediates of the TCA cycle, such as aKG, as biosynthetic precursors, and upregulate anaplerotic reactions that drive glutamine-derived carbon into the TCA cycle, counterbalancing biosynthetic efflux<sup>16</sup> (Supplementary Discussion). We hypothesized that in cells with high PHGDH expression, the PSAT1 reaction might contribute a significant fraction of glutamate to aKG flux. If true, the serine biosynthesis pathway would have an important role in TCA anaplerosis of glutamine-derived carbon. Consistent with this possibility, suppression of PHGDH in MDA-MB-468 cells caused a large reduction in the levels of aKG (Fig. 4c and Supplementary Fig. 7d). In fact, of the major metabolites measured, aKG was the one with the most significant and largest change upon PHGDH suppression, whereas serine levels were not significantly changed (Supplementary Fig. 8). PHGDH suppression also caused a significant reduction in other TCA components (Fig. 4d and Supplementary Fig. 8). Like suppression of PHGDH, suppression of PSAT1 also caused a significant reduction in serine pathway flux and aKG levels (Fig. 4c and Supplementary Fig. 7d, e). Furthermore, labelling studies using U-<sup>13</sup>C-glutamine revealed that the absolute flux from glutamine to aKG and other TCA intermediates was significantly reduced in cells with RNAi-mediated suppression of PHGDH or

PSAT1 (Fig. 4e and Supplementary Fig. 9a, b). These data indicate that in cell lines with high PHGDH expression, the serine synthesis pathway is responsible for approximately 50% of the net conversion of glutamate to aKG and that suppression of PHGDH results in a significant loss of TCA intermediate flux and steady-state levels of TCA intermediates (Fig. 4f and Supplementary Fig. 9a, b). Furthermore, labelling studies using U-<sup>13</sup>C-glucose in cell lines with *PHGDH* amplification (MDA-MB-468) and without (MDA-MB-231) revealed that in cells with high PHGDH expression, flux through the serine biosynthesis pathway shunts 8–9% of the glycolytic flux towards serine production, compared to 1–2% in the cell line with low PHGDH expression (Fig. 4f and Supplementary Fig. 9a). Therefore, increased flux through the serine biosynthesis pathway has a major impact on aKG production, but a smaller effect on glycolysis or serine availability in these cells (Supplementary Discussion). In contrast, another prominent aKG-producing transaminase, alanine aminotransferase, does not contribute significantly to aKG production in *PHGDH*-amplified cells (Supplementary Fig. 10).

We find that PHGDH expression is a critical part of a cellular program promoting serine pathway flux (Supplementary Fig. 5) and is responsible for a considerable portion of anaplerosis of glutamate into the TCA cycle as aKG (Supplementary Fig. 1). As ~70% of ER-negative breast cancers exhibit elevated PHGDH (Fig. 2c), our work suggests that targeting the serine synthesis pathway may be therapeutically valuable in breast cancers with elevated PHGDH expression or *PHGDH* amplifications (Supplementary Discussion). Lastly, we anticipate that the screening approach described here may be applicable to other cancer types or gene sets, enabling the identification of novel cancer targets directly in an *in vivo* context.

## METHODS SUMMARY

To undertake negative-selection RNAi screening in solid tumours, pools of MCF10DCIS.com cells expressing an shRNA library were injected into the fourth mammary fat pad of immunocompromised mice and allowed to form tumours. Abundances of shRNAs in the tumours were determined using massively parallel sequencing and compared to shRNA abundance in the injected cells. Genes targeted by shRNAs that were significantly depleted during tumour growth were considered hits and prioritized by analysing gene copy number data from human tumours and cancer cell lines. Lentiviral shRNAs were used to suppress PHGDH expression in breast cancer cell lines with and without *PHGDH* genomic amplification. Serine synthesis pathway activity and anaplerosis were measured via flux analyses using isotopically labelled molecules.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 October 2010; accepted 8 July 2011.

Published online 14 July 2011.

1. Jones, R. G. & Thompson, C. B. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev.* **23**, 537–548 (2009).

2. Hsu, P. P. & Sabatini, D. M. Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703–707 (2008).
3. Bric, A. *et al.* Functional identification of tumor-suppressor genes through an *in vivo* RNA interference screen in a mouse lymphoma model. *Cancer Cell* **16**, 324–335 (2009).
4. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
5. Miller, F. R., Santner, S. J., Tait, L. & Dawson, P. J. MCF10DCIS.com xenograft model of human comedo ductal carcinoma *in situ*. *J. Natl. Cancer Inst.* **92**, 1185–1186 (2000).
6. Beroukhi, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
7. Pollari, S. *et al.* Enhanced serine production by bone metastatic breast cancer cells stimulates osteoclastogenesis. *Breast Cancer Res. Treat.* **125**, 421–430 (2011).
8. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
9. Miller, F. R. *et al.* Xenograft model of progressive human proliferative breast disease. *J. Natl. Cancer Inst.* **85**, 1725–1732 (1993).
10. Snell, K. Enzymes of serine metabolism in normal, developing and neoplastic rat tissues. *Adv. Enzyme Regul.* **22**, 325–400 (1984).
11. Achouri, Y., Rider, M. H., Schaftingen, E. V. & Robbi, M. Cloning, sequencing and expression of rat liver 3-phosphoglycerate dehydrogenase. *Biochem. J.* **323**, 365–370 (1997).
12. Walsh, D. A. & Sallach, H. J. Purification and properties of chicken liver D-3-phosphoglycerate dehydrogenase. *Biochemistry* **4**, 1076–1085 (1965).
13. Knox, W. E., Herzfeld, A. & Hudson, J. Phosphoserine phosphatase distribution in normal and neoplastic rat tissues. *Arch. Biochem. Biophys.* **132**, 397–403 (1969).
14. Lund, K., Merrill, D. K. & Guynn, R. W. The reactions of the phosphorylated pathway of L-serine biosynthesis: thermodynamic relationships in rabbit liver *in vivo*. *Arch. Biochem. Biophys.* **237**, 186–196 (1985).
15. Wiederschain, D. *et al.* Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle* **8**, 498–504 (2009).
16. DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl Acad. Sci. USA* **104**, 19345–19350 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the D.M.S. laboratory, M. Kwon, B. Luo and F. Reinhardt for assistance. This research is supported by fellowships from S. G. Komen for the Cure to R.P. and the Life Science Research Foundation to Y.D.S. and grants from the Keck Foundation, the David H. Koch Institute for Integrative Cancer Research at MIT, The Alexander and Margaret Stewart Trust Fund, and NIH Grant CA103866 to D.M.S. D.M.S. is an investigator of the Howard Hughes Medical Institute.

**Author Contributions** R.P. and D.M.S. conceived the project and designed the experiments and K.M.M. and E.M.D. designed metabolic analysis and flux experiments. R.P. performed the screening, knockdown, cell proliferation and tumour formation experiments. K.B., Y.D.S., F.G.B., M.E.P., P.P.H., K.O., Z.-Y.T., N.Y.K. and W.W.C. assisted with experiments. E.F.B. collected breast cancer patient samples and scored PHGDH immunohistochemistry. D.K. performed and M.M.-K. assisted with interpretation of PHGDH immunohistochemistry. K.M.M., S.S., H.-K.W., A.K.J. and E.M.D. performed metabolite profiling, flux experiments and metabolite modelling. N.S., J.B. and L.A.G. assisted with amplification data. G.S.C. and D.E.R. assisted with screening technology. B.Y. and A.M.C. provided bioinformatic support. R.P. wrote and D.M.S. edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.M.S. ([sabatini@wi.mit.edu](mailto:sabatini@wi.mit.edu)).

## METHODS

**Materials.** Materials were obtained from the following sources: antibodies to PHGDH (HPA021241) and PSPH (HPA020376) from Sigma; an antibody against PYCR1 (13108-1-AP) from Proteintech; an antibody against GMPS (A302-417A) from Bethyl Labs; an antibody against VDACL1 (ab16814) from Abcam; antibodies to RPS6 (2217), PARP (9532) and Caspase-3 (9662) from Cell Signaling Technologies; an antibody against PSAT1 (H00029968-A01) from Novus Biologicals; an antibody against SLC16A3 (AB3316P) from Millipore; and HRP-conjugated anti-mouse, anti-rabbit secondary antibodies from Santa-Cruz Biotechnology; lactate dehydrogenase from Roche (10127230001); lactic acid from Acros; RPMI-1640 media, 3-bromopyruvate and glycine/hydrazine solution (G5418) from Sigma;  $\alpha$ - $^{15}\text{N}$ -glutamine from Isotech/Sigma (486809); L-[3H(G)]-serine from Perkin Elmer; Infinity Glucose Oxidase Liquid Stable Reagent (TR15221) from Thermo Electron; U- $^{13}\text{C}$ -glutamine from Isotech/Sigma (605166); MT-3 cells from DSMZ; Hs578T, MDA-MB-468, MDA-MB-231, BT-20, HCC1599, HCC70, DU4475, MCF-7 and ZR-75-30 cells from ATCC; MCF-10A, MCF-10AT1 and MCF10DCIS.com cells from the Karmanos Cancer Center, Michigan; matrigel from BD Biosciences; Phusion DNA polymerase from New England Biolabs; BCA Protein Assay from Pierce; siRNAs from Dharmacon; and amino-acid-free, glucose-free RPMI-1640 from US Biological. Lentiviral shRNAs were obtained from the The RNAi Consortium (TRC) collection of the Broad Institute<sup>4</sup>. The TRC numbers for the shRNAs used are: GFP, TRCN0000072186; PHGDH\_1, TRCN0000221861; PHGDH\_2, TRCN0000221865; PSPH\_1, TRCN0000002796; PSPH\_2, TRCN0000315168; PSAT1\_1, TRCN0000035266; PSAT1\_2, TRCN0000035268; SLC16A3\_1, TRCN0000038477; SLC16A3\_2, TRCN0000038478; VDACL1\_1, TRCN000029126; VDACL1\_2, TRCN000029127; GMPS\_1, TRCN0000045938; GMPS\_2, TRCN0000045941; PYCR1\_1, TRCN0000038979; PYCR1\_2, TRCN0000038980. The TRC website is <http://www.broadinstitute.org/rnai/trc/lib>. The doxycycline-inducible shRNA vector used was previously described<sup>15</sup>.

**Cell culture.** MDA-MB-468, MDA-MB-231, BT-20, HCC1599, HCC70, DU4475, ZR-75-30, MT-3, Hs578T and MCF-7 were cultured in RPMI supplemented with 10% FCS and penicillin/streptomycin. MCF-10A and MCF10AT1 cells were cultured as described previously<sup>17</sup>. MCF10DCIS.com cells were cultured in 50:50 DMEM and F12 media with 5% horse serum and penicillin/streptomycin.

**Compilation of metabolic gene list.** A list of all human metabolic enzymes and small molecule transporters was generated by cross-referencing maps of metabolic pathways (Roche) with the KEGG database (<http://www.genome.jp/kegg/kegg1.html>). NCBI resources including Entrez Gene (<http://www.ncbi.nlm.nih.gov/gene>) and the available literature were used to identify known or putative gene function and to identify functional homologues. A gene was considered a metabolic enzyme if it modified a small molecule to generate another small molecule. Genes which modified polymerized DNA or RNA or which modified proteins were excluded. In cases where an enzyme could modify both a small molecule and a macromolecule, we favoured a more liberal criterion of inclusion. A gene was considered a small molecule transporter if it formed a pore or channel through which a small molecule could traverse a lipid bilayer. Accessory or regulatory subunits of larger protein complexes were generally excluded.

**Meta-analysis of oncogenomic data.** To generate a cancer-relevant 'high priority' subset of metabolic genes (out of the 2,752 genes we classified as metabolic enzymes or small molecule transporters), we first identified those genes whose expression is significantly associated with the transformed state, advanced breast cancer, or stemness. Genes associated with the transformed state were obtained by analysing 36 gene expression studies deposited in Oncomine<sup>18</sup> that profiled normal human tissue and primary tumours derived from them. The gene expression profiles in each study were classified as normal or tumour and for each group the log<sub>2</sub> median centred intensity for each gene was determined. A *P* value associated with the significance of the difference between the two groups was calculated with the Student's *t*-test. After ranking the genes based on the *P* values, the top 10% of the genes with lowest *P* values were selected from each of the 36 studies. From these genes we identified those that are in the top 10% of the most upregulated metabolic genes across the all 36 studies at a *P* value < 0.05. Genes associated with aggressive breast cancer were obtained by analysing 15 gene expression studies from Oncomine that profiled ER<sup>-</sup> versus ER<sup>+</sup> tumours, grade 3 versus grade 1 or 2 tumours, tumours of basal versus epithelial morphology, or tumours from patients who failed to survive after 5 years of follow-up versus those who did survive at 5 years. The 15 studies were analysed as above to identify those genes that are in the top 10% of the most upregulated metabolic genes across the studies at a *P* value < 0.05. To identify genes associated with stemness, we analysed gene expression studies comparing differentiated cells with stem cells<sup>19</sup>, chromatin immunoprecipitation studies of stem-cell-associated transcription factors<sup>20,21</sup>, and a previous meta-analysis of stemness-associated genes<sup>22</sup>. Genes were considered to be associated with stemness if their average expression was greater than fourfold upregulated in the stem versus differentiated cells profiles analysed previously<sup>19</sup> or if

their promoters were bound by at least two stem-cell-specific transcription factors (Oct4, Nanog, Sox2, Tcf3, Dax1, Nac1 or Klf4) in both studies analysed. To generate the final high priority set of 133 genes that was screened (Supplementary Table 2), three categories of genes were selected: (1) genes scoring in all three analyses; (2) the most significantly scoring ~5% of genes in any one category; and (3) the most significantly scoring ~10% of genes in any two categories.

**Identification of cell lines for use in pooled screening.** To undertake negative-selection RNAi screening, a cell line that could form a tumour upon injection of the minimum number of cells was identified. To accomplish this, 11 breast cell lines that previously identified as capable of forming tumours were selected and 100,000 cells from each were injected into the fourth murine mammary fat pad. The cell lines tested included BT-20, BT-474, MCF10DCIS.com, HBL100, MCF7, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-453, T47D and ZR-75-1. After one month, tumours were scored by size and number scoring per site, and tumours or injection sites were analysed histologically to verify the presence of a tumour, or to identify microscopic tumours. In the timeframe of the experiment, MDA-MB-231, MDA-MB-361, MDA-MB-453, MCF7 and T47D cells formed microscopic tumours, whereas MCF10DCIS.com formed large tumours and ZR-75-1 formed small macroscopic tumours reproducibly. MCF10DCIS.com cells were then injected into murine mammary fat pads at 100,000, 10,000, 1,000 and 100 cells per site. All of these injections were capable of forming tumours, and tumour size correlated with the number of cells injected. The MCF10DCIS.com cell line was finally shown to be suitable for *in vivo* screening upon performing a screen using 180 shRNAs and demonstrating that nearly all of the shRNAs introduced initially could be recovered from the tumour and that replicate tumours exhibited significant correlation in those shRNAs over- or underrepresented compared to the injected pool. These experiments should not be construed to indicate that the excluded cell lines would not also be suitable for *in vivo* screening, as they were not tested using an shRNA pool.

**Pooled shRNA screening.** pLKO.1 lentiviral plasmids encoding shRNAs targeting the 133 transporters and metabolic enzymes listed in Supplementary Table 2 were obtained and combined to generate two plasmid pools. One contained the plasmids encoding shRNAs targeting all 47 transporters and another the plasmids encoding shRNAs targeting all 86 metabolic enzymes as well as control shRNAs designed not to target any gene. These plasmid pools were used to generate lentivirus-containing supernatants as described<sup>23</sup>. MCF10DCIS.com cells were infected with the pooled virus so as to ensure that each cell contained only one viral integrant. Cells were selected for 3 days with 0.5  $\mu\text{g ml}^{-1}$  puromycin. For the *in vivo* screen, cells were injected in 33% growth factor reduced matrigel into the fourth mammary fat pad of NOD.CB17 Scid/J mice (Jackson Labs) at 100,000 to 1,000,000 cells per injection site and tumours were harvested 4 weeks after implantation. For the *in vitro* screen, cells were plated in replicates of four at 1,000,000 per 10-cm plate and split at 1:8 once confluent (every 3–5 days) for 25–28 days. Genomic DNA was isolated from tumours or cells by digestion with proteinase K followed by isopropanol precipitation. To amplify the shRNAs encoded in the genomic DNA, PCR was performed for 33 cycles at an annealing temperature of 66 °C using 2–6  $\mu\text{g}$  of genomic DNA, the primer pair indicated below, and DNA polymerase. So that PCR products obtained from many different tumours could be sequenced together, forward primers containing unique 2-nucleotide barcodes were used (see below). After purification, the PCR products from each tumour were quantified by ethidium bromide staining after gel electrophoresis, pooled at equal proportions, and analysed by high-throughput sequencing (Illumina) using the primer indicated below. shRNAs from up to 16 genomic DNA samples were sequenced together. Sequencing reads were deconvoluted using GNU Octave software by segregating the sequencing data by barcode and matching the shRNA stem sequences to those expected to be present in the shRNA pool, allowing for mismatches of up to 3 nucleotides. The Log<sub>2</sub> values reported are the average Log base 2 of the fold change in the abundance of each shRNA in the pre-injection cells compared to tumours, for *n* = 5 tumours for the transporter pool and *n* = 12 tumours for the metabolic enzyme pool, or to cells at day 25–28 for *n* = 4 *in vitro* cultures. *P* values were determined by two-sided homoscedastic unpaired *t*-test comparing each shRNA to a basket of negative-control shRNAs contained within the shRNA pools. Individual shRNAs were identified as scoring in the screens using a *P* value cutoff of 0.05 and Log<sub>2</sub> fold-change cutoff of −1. Genes for which >75% of the shRNAs targeting the gene scored were considered hits. Individual shRNAs were considered to be differentially required *in vitro* versus *in vivo* using a *P*-value cutoff of 0.05 by a two-sided homoscedastic unpaired *t*-test comparing the *in vitro* and *in vivo* shRNA Log<sub>2</sub> fold change scores. For the transporter pool screen, this required normalization to the median of the two distributions. shRNAs present at less than 30 reads in the pre-injection cell sample were eliminated from further analysis.

Follow-up tumour growth studies of individual genes followed a similar timeline as above, except that during PHGDH and PSPH follow-up (Fig. 1e and



Supplementary Fig. 6), 10 days elapsed between infection and injections, whereas 5 days elapsed for all other validated genes (Supplementary Fig. 2). For doxycycline-inducible constructs, MDA-MB-468 cells were infected with *GFP*- or *PHGDH*-targeting shRNAs, puromycin selected and injected into the fourth murine mammary fat pad as above. Once tumours were palpable in all animals (25 days post-injection), doxycycline chow (600 p.p.m.) was provided to a randomly assigned set of animals for the duration of the experiment. Caliper measurements were taken every 4–6 days and tumour volume was estimated by  $0.5 \times W \times W \times L$ , where  $W$  is width and  $L$  is length. All experiments involving mice were carried out with approval from the Committee for Animal Care at MIT and under supervision of the Department of Comparative Medicine at MIT.

**Primers for amplifying shRNAs encoded in genomic DNA:** Barcoded forward primer ( $N$  indicates location of sample-specific barcode sequence): AATGATACGGCGACACCGAGAAAGTATTTCTGATTTCTTGGCTTTATATATCTTGTGGAANNACGAAAC. Common reverse primer: CAAGCAGAAGACGGCATA CGAGCTCTCCGATCTTGTGGATGAATACTGCCATTTGTCTCGAGGTC. Illumina sequencing primer: AGTATTTCTGATTTCTTGGCTTTATATATCTTGTGGAA.

**Analysis of gene copy number data.** The significance of copy number alteration across multiple data sets was determined using the GISTIC algorithm with methods described in ref. 6 and using the data deposited at <http://www.broadinstitute.org/tumourscape>.

**Determination of proportion of tumours with *PHGDH* overexpression.** To determine the percentage of breast cancers with elevations in *PHGDH* mRNA levels, data deposited in OncoPrint from ref. 8 were used. An ER<sup>+</sup> tumour was considered to have elevated *PHGDH* mRNA if the expression level was higher than 1.5 s.d. above the mean expression level in the ER<sup>+</sup> class (~91st percentile). For the percentage of breast cancer exhibiting elevated *PHGDH* protein, data reported in Fig. 2c were used. An ER<sup>+</sup> tumour was considered to have elevated *PHGDH* protein if the immunohistochemical staining signal was classified as 'high'.

**Cell proliferation assays.** For *PHGDH* or *PSPH* knockdown experiments, 10,000–20,000 MDA-MB-468, BT-20, HCC70, MCF-7, or MDA-MB-231 cells were infected with shRNA-expressing lentiviruses of known titres at a multiplicity of infection of 2.5 to 5. Cells were cultured in 12-well plates and infected via a 30-min spin at 2,250 r.p.m. in a Beckman Coulter Allegra X-12R centrifuge with an SX4750 rotor and uPlate Carrier attachment followed by an overnight incubation in media containing polybrene. Eight days after infection the number of cells was determined using a Coulter Counter (Beckman) and used to calculate relative cell proliferation. Where indicated, standard RPMI media was supplemented with serine to concentrations fivefold that of the serine already in the media. Where indicated, supplementation occurred at one and four days after lentiviral infection. For serine depletion experiments, cells were plated out as described above and the following day the standard culture medium was replaced with medium lacking serine or reconstituted with  $1 \times$  serine. Dialysed serum (3 kDa MWCO) was used in serine depletion experiments except in the case of MCF-10A cells, where standard 5% serum was used.

**Immunohistochemistry and immunoblotting.** Immunoblotting was performed as described<sup>24</sup>. *PHGDH* protein levels were quantified using an Odyssey Infrared Imager (Li-Cor). For each measurement, the *PHGDH* signal obtained was normalized to the RPS6 signal from the same lane after accounting for background fluorescence. Immunohistochemistry was performed on formalin-fixed paraffin-embedded sections using a boiling Dako antigen retrieval method, as described<sup>25</sup>. A 1:250 dilution of the *PHGDH* antibody was used. A pathologist scored, in a blinded fashion, the intensity of the *PHGDH* staining in the breast tumour samples using a scale of 0–3 that represents none/weak, moderate and strong staining. Use of the tumour samples for *PHGDH* staining was approved by Institutional Review Boards at MIT (Protocol Number 1005003872) and Massachusetts General Hospital (Protocol Number 2010-P-001505/1).

**Glucose and lactate measurements.** Cells infected with shRNAs were plated on the day after infection at 5,000 cells per well of a 96-well plate in RPMI-10 alone or with 25  $\mu$ M 3-bromopyruvate in a total of 200  $\mu$ l media. On day 4 after infection, media was collected from the wells and cells were washed once with phosphate buffered saline before lysis in 50 mM NaOH. Lysate was mixed well and protein measured by BCA protein assay (Pierce). To determine the integrated protein content over the course of the assay ( $\mu$ g protein  $\times$  days), a model was constructed with the following assumptions: control cells underwent two population doublings, cells proliferated exponentially to the final protein content, and the initial protein content for all samples was equivalent. Glucose concentration in the media was measured by glucose oxidase and peroxidase assay (Thermo Electron) and compared to control wells containing media with no cells to determine the quantity consumed. Lactate was measured by adding 5  $\mu$ l of media to a solution containing 0.3 M glycine/hydrazine solution (Sigma G5418), 2.4 mM NAD<sup>+</sup> (Fisher Scientific NC9877003), and 2  $\mu$ l ml<sup>-1</sup> lactate dehydrogenase (5 U  $\mu$ l<sup>-1</sup>,

Roche 10127230001) in a 200  $\mu$ l total volume in a 96-well microtitre plate. Plates were mixed briefly and incubated for 30 min at 37 °C before reading absorbance at 340 nm. Lactate concentration was determined by comparison to a lactic acid standard (10 mM–0 mM, Fisher Scientific AC18987-0050) and compared to control wells containing media with no cells to determine the quantity produced.

**Metabolite measurements.** For metabolite measurements, cells were cultured in cell-line-appropriate culture media (see above) in 10-cm dishes to approximately 70% confluence, typically by plating at  $2 \times 10^6$  cells per dish approximately 48 h before metabolite extraction. Twenty-four hours before metabolite extraction, culture media was replenished with media containing dialysed FBS. For metabolite extraction, cells in the culture dish were rapidly washed three times with 37 °C PBS, and then metabolites were extracted by addition of 80% aqueous methanol (pre-cooled in dry-ice) followed by incubation of culture dishes on dry ice for 15 min. For quantification, a <sup>13</sup>C-labelled internal metabolite standard for each analysed species was included in the extraction process. Cellular metabolite extracts were then collected by cell scraping and removal of the supernatant following centrifugation at 3,750 r.p.m. for 30 min (4 °C). The supernatants were then dried down using N<sub>2</sub> gas and stored dry at –80 °C before mass spectrometry analysis. Four biological replicate samples were generated and analysed for each cell line. In addition, two parallel dishes of cells were trypsinized and counted using a Nexcelom cell counter; subsequent metabolite measurements were normalized to cell count.

All cell extracts were analysed by liquid chromatography-triple quadrupole mass spectrometry (LC-MS) using scheduled selective reaction monitoring (SRM) for each metabolite of interest, with the detector set to negative mode. Prior to injection, dried extracts were reconstituted in LCMS grade water. LC separation was achieved by the method reported<sup>26</sup>. Extracted metabolite concentrations were calculated from standard metabolite build-up curves using natural <sup>12</sup>C synthetic metabolites and normalized against cell number as well as the internal <sup>13</sup>C-labelled metabolite standards added at the time of metabolite extraction.

**Flux analysis.** For aKG flux studies, cells were plated at 250,000 cells per well in 6-well culture dishes in typical culture media (see above). Twenty-four hours before the flux study timecourse, media was replenished with fresh RPMI media containing dialysed FBS. For the flux study timecourse, standard RPMI culture media with dialysed FBS was used and the glutamine was replaced with U-<sup>13</sup>C glutamine (2 mM final concentration, matching the glutamine concentration in standard RPMI culture media). At the relevant time points, metabolites were harvested as noted above.

Serine pathway flux was measured using extracellular  $\alpha$ -<sup>15</sup>N-glutamine, which is taken up by cells and becomes intracellular  $\alpha$ -<sup>15</sup>N-glutamate at a very high rate. The activity of PSAT1 (conversion of phospho-hydroxypyruvate to phosphoserine) is coupled to the transfer of the  $\alpha$ -<sup>15</sup>N-amino nitrogen of glutamate to phospho-hydroxypyruvate, generating aKG and  $\alpha$ -<sup>15</sup>N-phosphoserine. As extracellular serine is in equilibrium with the intracellular pool, the rate of accumulation of extracellular  $\alpha$ -<sup>15</sup>N-serine can be used to assess the activity of the serine biosynthetic pathway, and is proportional to the overall serine biosynthetic flux. For these flux studies, cells were plated at 250,000 cells per well in 6-well culture dishes in typical culture media (see above). When cells reached 60–70% confluence (typically 24–48 h post-cell plating), media was replenished with fresh media containing dialysed FBS and  $\alpha$ -<sup>15</sup>N-glutamine (2 mM final concentration). For the data presented in Fig. 3b, MCF10A medium was used (see above) to permit the inclusion of the MCF10A cell line, whereas for the data presented in Supplementary Fig. 6a, RPMI medium was used. Therefore, these data are not directly comparable between these two panels. Samples of media were collected from four biological replicates, at this initial time point and following 24 h of additional culture.  $\alpha$ -<sup>15</sup>N-serine was extracted from 300  $\mu$ l of the sample media by addition of 3 volumes of acetonitrile, followed by collection of the supernatant following centrifugation for 30 min at 3,750 r.p.m. Supernatant was then dried down using N<sub>2</sub> gas and the dry samples stored at –80 °C until mass spectrometry. In parallel with the metabolite extracts, two replicate wells were trypsinized and counted at the initial time point as well as the 24 h time point. The average of these four wells was used for subsequent cell number normalization. Prior studies established the linearity of production of serine over this timecourse, and demonstrated that the intracellular and extracellular serine pools are at steady-state equilibrium, enabling measurement of a lower-bound phosphoserine pathway flux by sampling extracellular  $\alpha$ -<sup>15</sup>N-serine. LC-MS analysis of <sup>15</sup>N-serine was performed using SRM in positive mode; separation was accomplished using an Atlantis HILIC Silica 5  $\mu$ m (2.1  $\times$  100 mm) column and a gradient of 10 mM of ammonium formate in Water (mobile phase A: aqueous 0.1% formic acid) and acetonitrile (mobile phase B, 0.1% formic acid) with mobile phase A linearly increasing from 5% to 60% over 4 minutes. Following a 2 min isocratic period, the system was returned to initial conditions for a total cycle time of 9 min at a flow rate of 200  $\mu$ l min<sup>-1</sup>. For flux studies, <sup>13</sup>C-labelled internal standards were omitted in both sample extracts and standard metabolite build-up curves.

**Flux modelling.** Ordinary differential equation models were constructed for two relevant portions of central carbon metabolism, based on the schematics shown in Supplementary Fig. 9a (models (i) and (ii)). Each model consisted of 3 differential equations with the constraints of balanced flux imposed on them. These equations describe the rates of loss of unlabelled forms of metabolites after feeding of 100% U-<sup>13</sup>C glucose or U-<sup>13</sup>C glutamine containing media.

The fluxes were identified by minimization of an objective function to the empirical data. The choice of objective function was  $\chi^2$ , defined as

$$\chi^2 = \sum_{k=1}^n \left( \frac{y_k - y(t_k; F)}{\sigma_k} \right)^2$$

where  $y_k$  is data point  $k$  with standard deviation  $\sigma_k$ , and  $y(t_k; F)$  is the value estimated by the model value at time point  $k$  for the set of fluxes  $F$ . Initial fluxes before the first optimization were arbitrarily chosen as 0.1. Three independent runs of 400 fits with the trust region approach were performed, each starting from the parameter values of the currently best fit randomly disturbed by up to 4 orders of magnitude.

**Model (i).** The schematic of the upper part of glycolysis (Supplementary Fig. 9a (i)) shows that  $F_2$  is the upper bound of the glycolytic flux that can be diverted to the pSer pathway. We estimated  $F_2$  by fitting the model to the time course of unlabelled metabolites (3PG, PEP and lactate) obtained using LC-MS of extracts from MDA-MB-468 and MDA-MB-231 cell lines, amplified and non-amplified PHGDH cell lines respectively. Three independent simulations of 400 fits were run for both the cell lines. The quality of fit was characterized by  $\chi^2$  values. The best 10% of the fits that also had  $P$  value above a significance threshold (0.05) were chosen for the analysis. The values of the parameter had a high variability suggesting that the parameter search space resembled a shallow basin. This was confirmed by generating the  $\chi^2$  landscapes for all possible pairs of parameters (data not shown). This observation suggested that additional constraints would greatly improve the predictive power of our model. Because each molecule of glucose that proceeds through glycolysis is broken into two molecules of 3PG, we imposed the requirement that  $F_1$  cannot be greater than twice the measured glucose consump-

tion rates (82 nmol per million cells per min). This additional constraint narrows down the solution of fluxes significantly, providing the results reported in the tables.

**Model (ii).** The schematic of the upstream reactions in glutaminolysis (Supplementary Fig. 9a (ii)) shows that  $F_1 + F_2$  is the glutamate to aKG flux. We estimated the fluxes as described above by fitting the model to the time course of unlabelled metabolites (glutamine, glutamate and aKG) obtained using LC-MS for MDA-MB468 cells with and without PHGDH suppression via RNAi. Identical statistical thresholds were applied as for model (i) (top 10% and  $P > 0.05$ ) to chose solutions for the analysis. Unlike model (i), the parameters converged very well without need for further constraint, confirmed by generating the  $\chi^2$  landscapes for all possible pairs of parameters (data not shown).

17. Debnath, J. & Brugge, J. S. Modelling glandular epithelial cancers in three-dimensional cultures. *Nature Rev. Cancer* **5**, 675–688 (2005).
18. Rhodes, D. R. *et al.* Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166–180 (2007).
19. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
20. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
21. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
22. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genet.* **40**, 499–507 (2008).
23. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 20380–20385 (2008).
24. Sancak, Y. *et al.* Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* **141**, 290–303 (2010).
25. Kalaany, N. Y. & Sabatini, D. M. Tumours with PI3K activation are resistant to dietary restriction. *Nature* **458**, 725–731 (2009).
26. Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh performance liquid chromatography–tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem.* **82**, 4403–4412 (2010).

# Crystal structure of the $\beta_2$ adrenergic receptor–Gs protein complex

Søren G. F. Rasmussen<sup>1,2\*</sup>, Brian T. DeVree<sup>3\*</sup>, Yaozhong Zou<sup>1</sup>, Andrew C. Kruse<sup>1</sup>, Ka Young Chung<sup>1</sup>, Tong Sun Kobilka<sup>1</sup>, Foon Sun Thian<sup>1</sup>, Pil Seok Chae<sup>4</sup>, Els Pardon<sup>5,6</sup>, Diane Calinski<sup>3</sup>, Jesper M. Mathiesen<sup>1</sup>, Syed T. A. Shah<sup>7</sup>, Joseph A. Lyons<sup>7</sup>, Martin Caffrey<sup>7</sup>, Samuel H. Gellman<sup>4</sup>, Jan Steyaert<sup>5,6</sup>, Georgios Skiniotis<sup>8</sup>, William I. Weiss<sup>1,9</sup>, Roger K. Sunahara<sup>3</sup> & Brian K. Kobilka<sup>1</sup>

G protein-coupled receptors (GPCRs) are responsible for the majority of cellular responses to hormones and neurotransmitters as well as the senses of sight, olfaction and taste. The paradigm of GPCR signalling is the activation of a heterotrimeric GTP binding protein (G protein) by an agonist-occupied receptor. The  $\beta_2$  adrenergic receptor ( $\beta_2$ AR) activation of Gs, the stimulatory G protein for adenylyl cyclase, has long been a model system for GPCR signalling. Here we present the crystal structure of the active state ternary complex composed of agonist-occupied monomeric  $\beta_2$ AR and nucleotide-free Gs heterotrimer. The principal interactions between the  $\beta_2$ AR and Gs involve the amino- and carboxy-terminal  $\alpha$ -helices of Gs, with conformational changes propagating to the nucleotide-binding pocket. The largest conformational changes in the  $\beta_2$ AR include a 14 Å outward movement at the cytoplasmic end of transmembrane segment 6 (TM6) and an  $\alpha$ -helical extension of the cytoplasmic end of TM5. The most surprising observation is a major displacement of the  $\alpha$ -helical domain of Gs relative to the Ras-like GTPase domain. This crystal structure represents the first high-resolution view of transmembrane signalling by a GPCR.

## Introduction

The  $\beta_2$  adrenergic receptor ( $\beta_2$ AR) has been a model system for the large and diverse family of G protein-coupled receptors (GPCRs) for over 40 years. It was one of the first GPCRs to be characterized by radioligand binding, and it was the first neurotransmitter receptor to be cloned<sup>1</sup> and structurally determined by crystallography<sup>2,3</sup>. The  $\beta_2$ AR was initially identified based on its physiological and pharmacological properties, but it was not known if receptors and G proteins were separate entities, or parts of the same protein<sup>4</sup>. Subsequent biochemical studies led to the isolation and purification of functional  $\beta_2$ AR and Gs, the stimulatory G protein that activates adenylyl cyclase, and the reconstitution of this signalling complex in phospholipid vesicles<sup>5,6</sup>. The cooperative interactions of  $\beta_2$ AR and Gs observed in ligand binding assays formed the foundation of the ternary complex model of GPCR activation<sup>7,8</sup>. In the ternary complex consisting of agonist, receptor and G protein, the affinity of the receptor for agonist is enhanced and the specificity of the G protein for guanine nucleotides changes in favour of GTP over GDP. The GPCR field has evolved markedly since these initial studies. Isolation of the genes and cDNAs for the  $\beta_2$ AR and other GPCRs using protein sequencing and expression cloning led to the expansion of the family by homology cloning. More recently, sequencing of the human genome led to the identification of over 800 GPCR genes<sup>9</sup>. Experimental tools for identifying protein–protein interactions and for expression and silencing of genes have revealed a complex network of cellular signalling and regulatory pathways including G protein-independent activation of cytosolic kinases<sup>10,11</sup>. Nevertheless, the  $\beta_2$ AR continues to be a relevant model for most aspects of GPCR pharmacology, signalling and regulation.

Notwithstanding the remarkable advances in this field, we still know relatively little about the structural basis for transmembrane signalling by GPCRs. Figure 1 shows the G protein cycle for the  $\beta_2$ AR–Gs complex. Agonist binding to the  $\beta_2$ AR promotes interactions with GDP-bound Gs $\alpha\beta\gamma$  heterotrimer, leading to the exchange of GDP for GTP, and the functional dissociation of Gs into G $\alpha$ -GTP and G $\beta\gamma$  subunits. The separate G $\alpha$ -GTP and G $\beta\gamma$  subunits can modulate the activity of different cellular effectors (channels, kinases or other enzymes). The intrinsic GTPase activity of G $\alpha$ s leads to hydrolysis of GTP to GDP and the reassociation of G $\alpha$ -GDP and G $\beta\gamma$  subunits, and the termination of signalling. The active state of a GPCR can be defined as that conformation that couples to and stabilizes a nucleotide-free G protein. In this agonist– $\beta_2$ AR–Gs ternary complex, Gs has a higher affinity for GTP than GDP, and the  $\beta_2$ AR has an approximately 100-fold higher affinity for agonists than does  $\beta_2$ AR alone. In an effort to understand the structural basis for GPCR signalling, we crystallized the  $\beta_2$ AR–Gs complex.

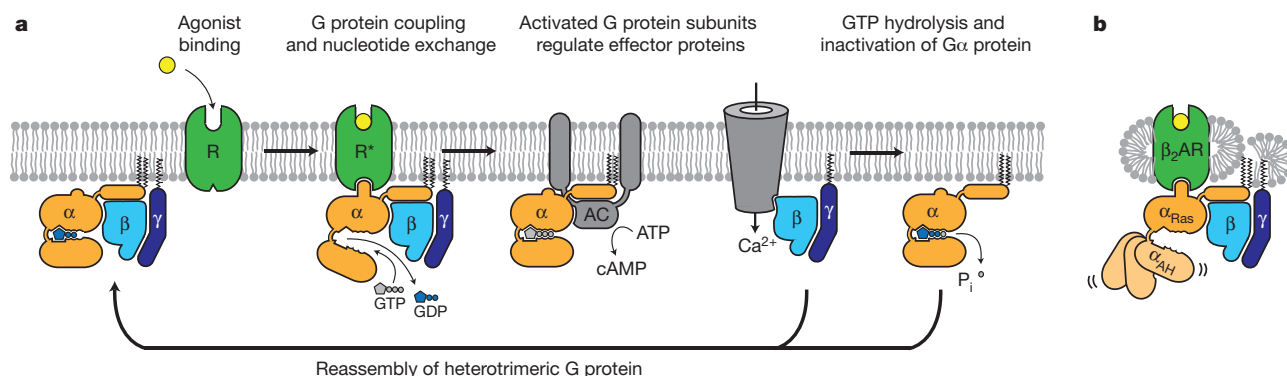
## Crystallization of the $\beta_2$ AR–Gs complex

The first challenge for crystallogenes was to prepare a stable  $\beta_2$ AR–Gs complex in detergent solution. The  $\beta_2$ AR and Gs couple efficiently in lipid bilayers, but not in detergents used to solubilize and purify these proteins. We found that a relatively stable  $\beta_2$ AR–Gs complex could be prepared by mixing purified GDP-Gs (approximately 100  $\mu$ M final concentration) with a molar excess of purified  $\beta_2$ AR bound to a high affinity agonist (BI-167107, Boehringer Ingelheim)<sup>12</sup> in dodecylmalto-side solution. Apyrase, a non-selective purine pyrophosphatase, was added to hydrolyze GDP released from Gs on forming a complex with the  $\beta_2$ AR. Removal of GDP was essential because both GDP and GTP

<sup>1</sup>Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Department of Neuroscience and Pharmacology, The Panum Institute, University of Copenhagen, 2200 Copenhagen N, Denmark. <sup>3</sup>Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. <sup>4</sup>Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>5</sup>Department of Molecular and Cellular Interactions, Vlaams Instituut voor Biotechnologie (VIB), Vrije Universiteit Brussel, B-1050 Brussel, Belgium. <sup>6</sup>Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium. <sup>7</sup>Membrane Structural and Functional Biology Group, Schools of Medicine and Biochemistry & Immunology, Trinity College, Dublin 2, Ireland. <sup>8</sup>Life Sciences Institute and Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>9</sup>Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

\*These authors contributed equally to this work.





**Figure 1 | G protein cycle for the  $\beta_2$ AR-Gs complex.** **a**, Extracellular agonist binding to the  $\beta_2$ AR leads to conformational rearrangements of the cytoplasmic ends of transmembrane segments that enable the Gs heterotrimer ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) to bind the receptor. GDP is released from the  $\alpha$  subunit upon formation of  $\beta_2$ AR-Gs complex. The GTP binds to the nucleotide-free  $\alpha$  subunit resulting in dissociation of the  $\alpha$  and  $\beta\gamma$  subunits from the receptor. The subunits regulate their respective effector proteins adenylyl cyclase (AC) and  $\text{Ca}^{2+}$  channels. The Gs heterotrimer reassembles from  $\alpha$  and  $\beta\gamma$  subunits following hydrolysis of GTP to GDP in the  $\alpha$  subunit. **b**, The purified nucleotide-free  $\beta_2$ AR-Gs protein complex maintained in detergent micelles. The G $\alpha$ s subunit consists of two domains, the Ras domain ( $\alpha$ Ras) and the  $\alpha$ -helical domain ( $\alpha$ AH). Both are involved in nucleotide binding. In the nucleotide-free state, the  $\alpha$ AH domain has a variable position relative the  $\alpha$ Ras domain.

can disrupt the high-affinity interaction between  $\beta_2$ AR and Gs (Supplementary Fig. 1a). The complex was subsequently purified by sequential antibody affinity chromatography and size-exclusion chromatography. The stability of the complex was enhanced by exchanging it into a recently developed maltose neopentyl glycol detergent (NG-310, Anatrace)<sup>13</sup>. The complex could be incubated at room temperature for 24 h without any noticeable degradation; however, initial efforts to crystallize the complex using sparse matrix screens in detergent micelles, bicelles and lipidic cubic phase (LCP) failed.

To further assess the quality of the complex, we analysed the protein by single particle electron microscopy. The results, which are described in detail in a companion manuscript (Westfield *et al.*, manuscript submitted), confirmed that the complex was monodisperse, but revealed two potential problems for obtaining diffraction of quality crystals. First, the detergent used to stabilize the complex formed a large micelle, leaving little polar surface on the extracellular side of the  $\beta_2$ AR-Gs complex for the formation of crystal lattice contacts. Our initial approach to this problem, which was to generate antibodies to the extracellular surface, was not successful. As an alternative approach, we replaced the unstructured amino terminus of the  $\beta_2$ AR with T4 lysozyme (T4L). We previously used T4L to facilitate crystallogensis of the inactive  $\beta_2$ AR by inserting T4L between the cytoplasmic ends of TM5 and TM6 (ref. 3). Several different amino-terminal fusion proteins were prepared and single particle electron microscopy was used to identify a fusion with a relatively fixed orientation of T4L in relation to the  $\beta_2$ AR.

The second problem revealed by single particle electron microscopy analysis was increased variability in the positioning of the  $\alpha$ -helical component of the G $\alpha$ s subunit. G $\alpha$ s consists of two domains, the Ras-like GTPase domain (G $\alpha$ sRas), which interacts with the  $\beta_2$ AR and the G $\beta$  subunit, and the  $\alpha$ -helical domain (G $\alpha$ sAH)<sup>14</sup>. The interface of the two G $\alpha$ s subdomains forms the nucleotide-binding pocket (Fig. 1), and electron microscopy two-dimensional (2D) averages and three-dimensional (3D) reconstructions show that in the absence of guanine nucleotide, G $\alpha$ sAH has a variable position relative to the complex of T4L- $\beta_2$ AR-G $\alpha$ sRas-G $\beta\gamma$  (Fig. 1b) (Westfield *et al.*, manuscript submitted).

We attributed the variable position of G $\alpha$ sAH to the empty nucleotide-binding pocket. However, as noted above both GDP and non-hydrolysable GTP analogues disrupt the  $\beta_2$ AR-Gs complex (Supplementary Fig. 1). The addition of pyrophosphate and its analogue phosphonoformate (foscarnet) led to a significant increase in stabilization of G $\alpha$ sAH as determined by electron microscopy analysis of the detergent-solubilized complex (Westfield *et al.*, manuscript submitted). Crystallization trials were carried out in lipidic cubic phase

and  $\text{Ca}^{2+}$  channels. The Gs heterotrimer reassembles from  $\alpha$  and  $\beta\gamma$  subunits following hydrolysis of GTP to GDP in the  $\alpha$  subunit. **b**, The purified nucleotide-free  $\beta_2$ AR-Gs protein complex maintained in detergent micelles. The G $\alpha$ s subunit consists of two domains, the Ras domain ( $\alpha$ Ras) and the  $\alpha$ -helical domain ( $\alpha$ AH). Both are involved in nucleotide binding. In the nucleotide-free state, the  $\alpha$ AH domain has a variable position relative the  $\alpha$ Ras domain.

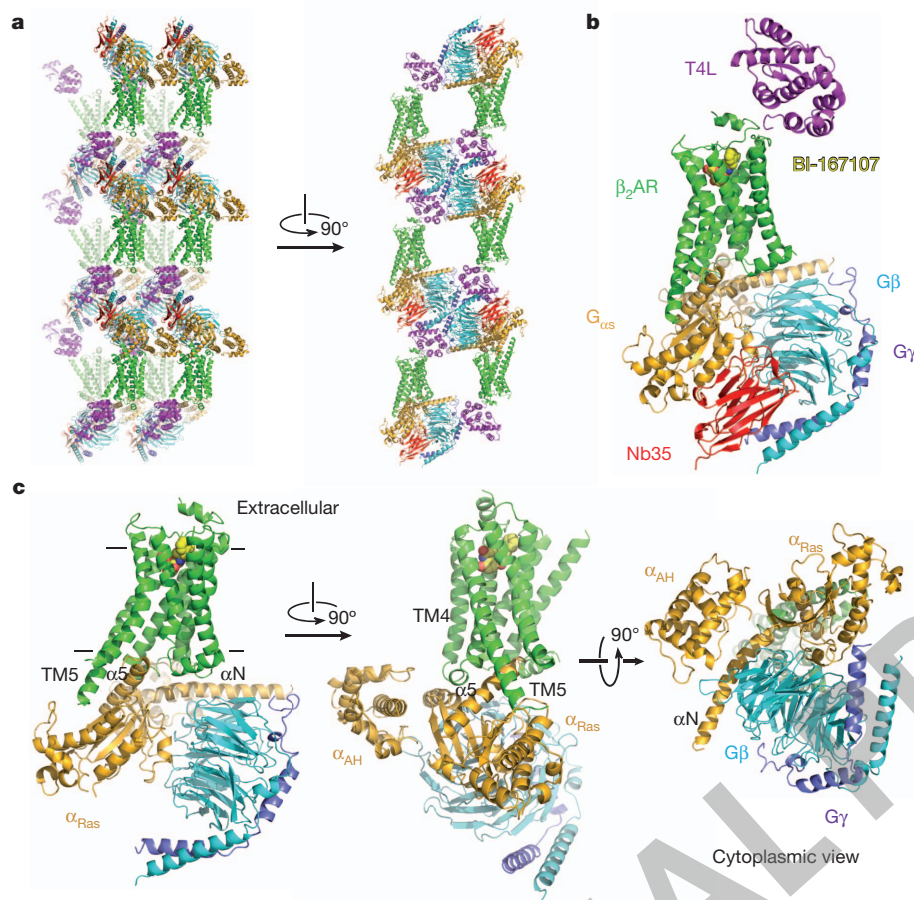
(LCP) using a modified monolein (7.7, see Methods) designed to accommodate the large hydrophilic component of the T4L- $\beta_2$ AR-Gs complex<sup>15</sup>. Although we were able to obtain small crystals that diffracted to 7 Å, we were unable to improve their quality through the use of additives and other modifications.

In an effort to generate an antibody that would further stabilize the complex and facilitate crystallogensis, we crosslinked  $\beta_2$ AR and the Gs heterotrimer with a small, homobifunctional amine-reactive cross-linker and used this stabilized complex to immunize llamas. Llamas and other camelids produce antibodies devoid of light chains. The single domain antigen binding fragments of these heavy-chain-only antibodies, known as nanobodies, are small (15 kDa), rigid and are easily cloned and expressed in *Escherichia coli* (Methods)<sup>16</sup>. We obtained a nanobody (Nb35) that binds to the complex and prevents dissociation of the complex by GTP $\gamma$ S (Supplementary Fig. 1). The T4L- $\beta_2$ AR-Gs-Nb35 complex was used to obtain crystals that grew to 250  $\mu\text{m}$  (Supplementary Fig. 2) in LCP (monolein 7.7) and diffracted to 2.9 Å. A 3.2 Å data set was obtained from 20 crystals and the structure was determined by molecular replacement (Methods).

The  $\beta_2$ AR-Gs complex crystallized in primitive monoclinic space group  $P2_1$ , with a single complex in each asymmetric unit. Figure 2a shows the crystallographic packing interactions. Complexes are arrayed in alternating aqueous and lipidic layers with lattice contacts formed almost exclusively between soluble components of the complex, leaving receptor molecules suspended between G protein layers and widely separated from one another in the plane of the membrane. Extensive lattice contacts are formed among all the soluble proteins, probably accounting for the strong overall diffraction and remarkably clear electron density for the G protein. Nb35 and T4L facilitated crystal formation. Nb35 packs at the interface of the G $\beta$  and G $\alpha$  subunits, with the complementarity determining region (CDR) 1 interacting primarily with G $\beta$  and a long CDR3 loop interacting with both G $\beta$  and G $\alpha$  subunits. The framework regions of Nb35 from one complex also interact with G $\alpha$  subunits from two adjacent complexes. T4L is linked to the  $\beta_2$ AR only through amino-terminal fusion, but packs against the amino terminus of the G $\beta$  subunit of one complex, the carboxy terminus of the G $\gamma$  subunit of another complex, and the G $\alpha$  subunit of yet another complex. Figure 2b shows the structure of the complete complex including T4L and Nb35, and Fig. 2c shows the  $\beta_2$ AR-Gs complex alone.

## Structure of the active-state $\beta_2$ AR

The  $\beta_2$ AR-Gs structure provides the first high-resolution insight into the mechanism of signal transduction across the plasma membrane



**Figure 2 | Overall structure of the  $\beta_2$ AR-Gs complex.** **a**, Lattice packing of the complex shows alternating layers of receptor and G protein within the crystal. Abundant contacts are formed among proteins within the aqueous layers. **b**, The overall structure of the asymmetric unit contents shows the  $\beta_2$ AR (green) bound to an agonist (yellow spheres) and engaged in extensive interactions with G $\alpha$ s (orange). G $\alpha$ s together with G $\beta$  (cyan) and G $\gamma$  (purple) constitute the heterotrimeric G protein Gs. A Gs-binding nanobody (red) binds the G protein between the  $\alpha$  and  $\beta$  subunits. The nanobody (Nb35) facilitates crystallization, as does T4 lysozyme (magenta) fused to the amino terminus of the  $\beta_2$ AR. **c**, The biological complex omitting crystallization aids, showing its location and orientation within a cell membrane.

by a GPCR, and the structural basis for the functional properties of the ternary complex. Figure 3a compares the structures of the agonist-bound receptor in the  $\beta_2$ AR-Gs complex and the inactive carazolol-bound  $\beta_2$ AR. The largest difference between the inactive and active structures is a 14 Å outward movement of TM6 when measured at the C $\alpha$  carbon of E268. There is a smaller outward movement and extension of the cytoplasmic end of the TM5 helix by 7 residues. A stretch of 26 amino acids in the third intracellular loop (ICL3) is disordered. Another notable difference between inactive and active structures is the second intracellular loop (ICL2), which forms an extended loop in the inactive  $\beta_2$ AR structure and an  $\alpha$ -helix in the  $\beta_2$ AR-Gs complex. This helix is also observed in the  $\beta_2$ AR-Nb80 structure (Fig. 3b); however, it may not be a feature that is unique to the active state, because it is also observed in the inactive structure of the highly homologous avian  $\beta_1$ AR (ref. 17).

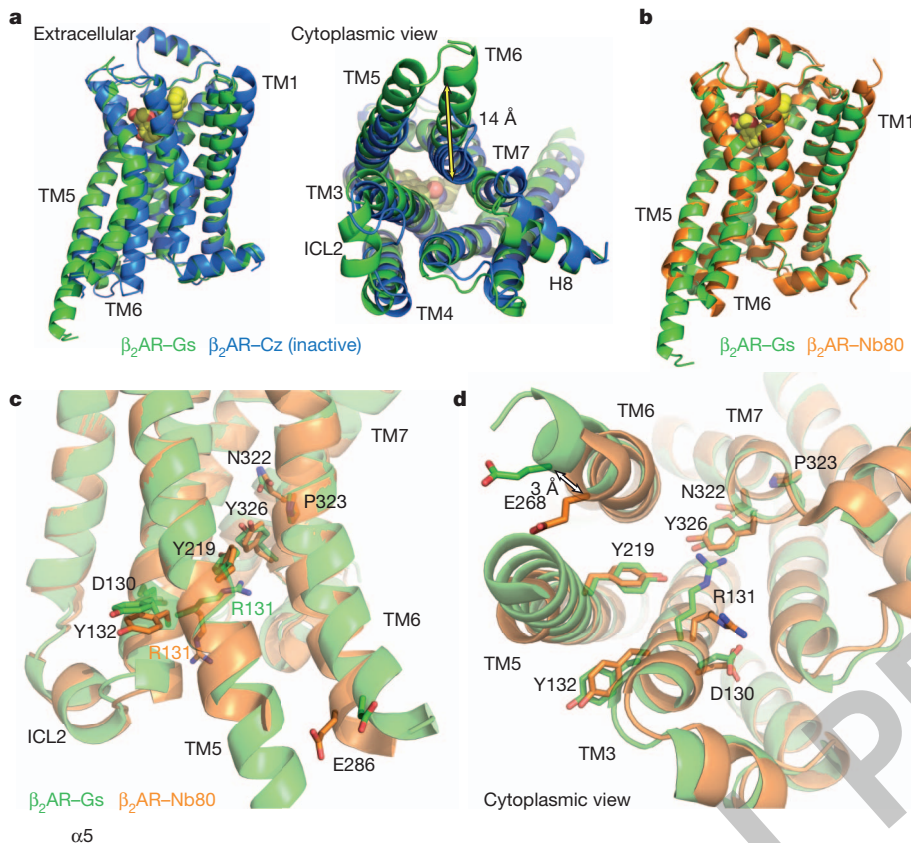
The quality of the electron density maps for the  $\beta_2$ AR is highest at the  $\beta_2$ AR-G $\alpha$ sRas interface, and much weaker for the extracellular half. The extracellular half of the receptor is not stabilized by any packing interactions either laterally with adjacent receptors in the membrane or through the extracellular surface. Instead, the extracellular region is indirectly tethered to the well-packed soluble components by the amino-terminal fusion to T4 lysozyme (Fig. 2a). Given the flexible and dynamic nature of GPCRs, the absence of stabilizing packing interactions may lead to structural heterogeneity in the extracellular half of the receptor and, consequently, to the limited quality of the electron density maps. However, the overall structure of the  $\beta_2$ AR in the T4L- $\beta_2$ AR-Gs complex is very similar to our recent active-state structure of  $\beta_2$ AR stabilized by a G protein mimetic nanobody (Nb80)<sup>12</sup>. In the  $\beta_2$ AR-Nb80 crystal, each receptor molecule has extensive packing interactions with adjacent receptors and the quality of the electron density maps for the agonist-bound  $\beta_2$ AR in this complex is remarkably good for a 3.5 Å structure. Therefore, the

$\beta_2$ AR-Nb80 structure allows us to confidently model BI-167107 here, and provide a more reliable view of the conformational rearrangements of amino acids around the ligand-binding pocket and between the ligand-binding pocket and the Gs-coupling interface<sup>12</sup>.

The overall root mean square deviation between the  $\beta_2$ AR components in the  $\beta_2$ AR-Gs and  $\beta_2$ AR-Nb80 structures is approximately 0.6 Å, and they differ most at the cytoplasmic ends of transmembrane helices 5 and 6 where they interact with the different proteins (Fig. 3b–d). The largest divergence is a 3 Å outward movement at the end of helix 6 in the  $\beta_2$ AR-Gs complex. However, the differences between these two structures are very small at the level of the most highly conserved amino acids (E/DRY and NPxxY), which are located at the cytoplasmic ends of the transmembrane segments (Fig. 3c,d). These conserved sequences have been proposed to be important for activation or for maintaining the receptor in the inactive state<sup>18</sup>. Of these residues, only Arg 131 differs significantly between these two structures. In  $\beta_2$ AR-Nb80 Arg 131 interacts with Nb80, whereas in the  $\beta_2$ AR-Gs structure Arg 131 packs against Tyr 391 of G $\alpha$ s (Supplementary Fig. 3). The high structural similarity is in agreement with the functional similarity of these two proteins. The  $\beta_2$ AR-Nb80 complex shows the same high affinity for the agonist isoproterenol as does the  $\beta_2$ AR-Gs complex<sup>12</sup>, consistent with high structural homology around the ligand binding pocket.

The active state of the  $\beta_2$ AR is stabilized by extensive interactions with G $\alpha$ sRas (Fig. 4). There are no direct interactions with G $\beta$  or G $\gamma$  subunits. The total buried surface of the  $\beta_2$ AR-G $\alpha$ sRas interface is 2,576 Å<sup>2</sup> (1,300 Å<sup>2</sup> for G $\alpha$ sRas and 1,276 Å<sup>2</sup> for the  $\beta_2$ AR). This interface is formed by ICL2, TM5 and TM6 of the  $\beta_2$ AR, and by  $\alpha$ 5-helix, the  $\alpha$ N- $\beta$ 1 junction, the top of the  $\beta$ 3-strand, and the  $\alpha$ 4-helix of G $\alpha$ sRas (see Supplementary Table 1 for specific interactions). Some of the  $\beta_2$ AR sequences involved in this interaction have been shown to have a role in G protein coupling; however, there is no clear consensus



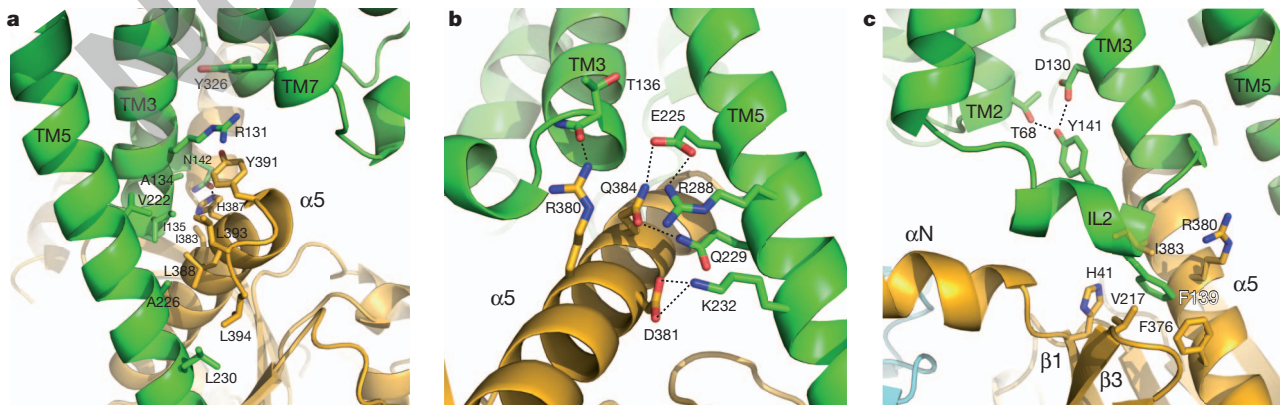


**Figure 3 | Comparison of active and inactive  $\beta_2\text{AR}$  structures.** **a**, Side and cytoplasmic views of the  $\beta_2\text{AR-Gs}$  structure (green) compared to the inactive carazolol-bound  $\beta_2\text{AR}$  structure<sup>3</sup> (blue). Significant structural changes are seen for the intracellular domains of TM5 and TM6. TM5 is extended by two helical turns whereas TM6 is moved outward by 14 Å as measured at the  $\alpha$ -carbons of Glu 268 (yellow arrow) in the two structures. **b**,  $\beta_2\text{AR-Gs}$  compared with the nanobody-stabilized active state  $\beta_2\text{AR-Nb80}$  structure<sup>12</sup> (orange). **c**, The positions of residues in the E/DRY and NPxxY motifs and other key residues of the  $\beta_2\text{AR-Gs}$  and  $\beta_2\text{AR-Nb80}$  structures. All residues occupy very similar positions except Arg 131 which in the  $\beta_2\text{AR-Nb80}$  structure interacts with the nanobody. **d**, View from the cytoplasmic side of residues shown in (c).

sequence for Gs-coupling specificity when these segments are aligned with other GPCRs. Perhaps this is not surprising considering that the  $\beta_2\text{AR}$  also couples to  $G_i$  and that many GPCRs couple to more than one G protein isoform. Of the 21 amino acids of Gs that are within 4 Å of the  $\beta_2\text{AR}$ , only five are identical between Gs and  $G_i$ , and all of these are in the carboxy-terminal  $\alpha$  helix. The structural basis for G protein coupling specificity must therefore involve more subtle features of the secondary and tertiary structure. Nevertheless, a noteworthy interaction involves Phe 139, which is located at the beginning of the ICL2 helix and sits in a hydrophobic pocket formed by Gs His 41 at the beginning of the  $\beta 1$ -strand, Val 217 at the start of the  $\beta 3$ -strand and Phe 376, Cys 379, Arg 380 and Ile 383 in the  $\alpha 5$ -helix (Fig. 4c). The

$\beta_2\text{AR}$  mutant F139A has severely impaired coupling to Gs<sup>19</sup>. The residue corresponding to Phe 139 is a Phe or Leu on almost all Gs coupled receptors, but is more variable in GPCRs known to couple to other G proteins. Of interest, the ICL2 helix is stabilized by an interaction between Asp 130 of the conserved DRY sequence and Tyr 141 in the middle of the ICL2 helix (Fig. 4c). Tyr 141 has been shown to be a substrate for the insulin receptor tyrosine kinase<sup>20</sup>; however, the functional significance of this phosphorylation is currently unknown.

The lack of direct interactions between the  $\beta_2\text{AR}$  and  $G\beta\gamma$  is somewhat unexpected given that a heterotrimer is required for efficient coupling to a GPCR. Whereas  $G\beta$  does not interact directly with the  $\beta_2\text{AR}$ , it has an indirect but important role in coupling by stabilizing



**Figure 4 | Receptor-G protein interactions.** **a**, **b**, The  $\alpha 5$ -helix of Gs docks into a cavity formed on the intracellular side of the receptor by the opening of transmembrane helices 5 and 6. **a**, Within the transmembrane core, the interactions are primarily non-polar. An exception involves packing of Tyr 391 of the  $\alpha 5$ -helix against Arg 131 of the conserved DRY sequence in TM3 (see also Supplementary Fig. 3). Arg 131 also packs against Tyr 326 of the conserved

NPxxY sequence in TM7. **b**, As  $\alpha 5$ -helix exits the receptor it forms a network of polar interactions with TM5 and TM3. **c**, Receptor residues Thr 68 and Asp 130 interact with the ICL2 helix of the  $\beta_2\text{AR}$  via Tyr 141, positioning the helix so that Phe 139 of the receptor docks into a hydrophobic pocket on the G protein surface, thereby structurally linking receptor-G protein interactions with the highly conserved DRY motif of the  $\beta_2\text{AR}$ .



the amino-terminal  $\alpha$  helix of  $G\alpha_s$  (Fig. 2C). Several models involving GPCR dimers propose that one of the protomers interacts predominantly with  $G\alpha$  while the other interacts with  $G\beta\gamma$ <sup>21–23</sup>. Consistent with these models, biochemical and biophysical evidence suggests that  $G\alpha_i2$  forms a stable complex with a LTB4 receptor dimer<sup>24</sup>. Whereas the  $\beta_2$ AR efficiently activates Gs as a monomer, extensive biochemical and biophysical evidence supports the existence of  $\beta_2$ AR dimers or oligomers in living cells<sup>25</sup>. Therefore, we cannot exclude the possibility that in cell membranes one protomer of a  $\beta_2$ AR dimer may interact with the  $G\beta\gamma$  subunit.

## Structure of activated Gs

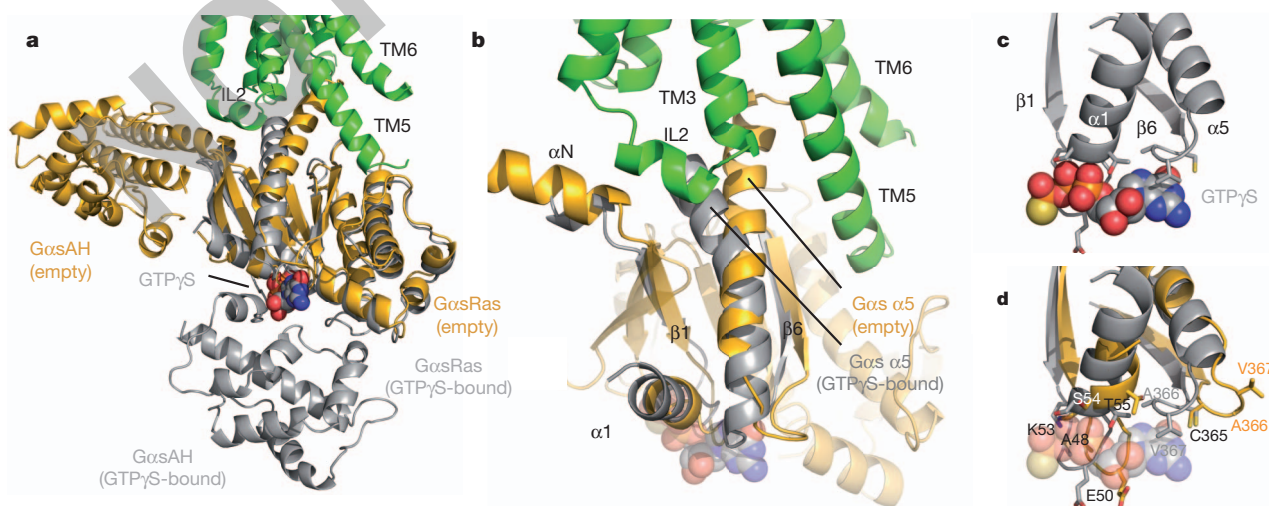
The most surprising observation in the  $\beta_2$ AR–Gs complex is the large displacement of the  $G\alpha_s$ AH relative to  $G\alpha_s$ Ras (an approximately 127° rotation about the junction between the domains) (Fig. 5a).  $G\alpha_s$ AH moves as a rigid body as shown by the alignment of  $\beta_2$ AR–Gs with  $G\alpha_s$ AH from the crystal structure of  $G\alpha_s$ –GTP $\gamma$ S<sup>26</sup> (Supplementary Fig. 4). In the structure of  $G\alpha_s$ –GTP $\gamma$ S, the nucleotide-binding pocket is formed by the interface between  $G\alpha_s$ Ras and  $G\alpha_s$ AH. Guanine nucleotide binding stabilizes the interaction between these two domains. The loss of this stabilizing effect of guanine nucleotide binding is consistent with the high flexibility observed for  $G\alpha_s$ AH in single particle electron microscopy analysis of the detergent-solubilized complex (Westfield *et al.*, manuscript submitted). It is also in agreement with the increase in deuterium exchange at the interface between these two domains upon formation of the complex (Chung *et al.*, manuscript submitted). Recently double electron-electron resonance (DEER) spectroscopy was used to document large (up to 20 Å) changes in distance between nitroxide probes positioned on the Ras and  $\alpha$ -helical domains of Gi upon formation of a complex with light-activated rhodopsin<sup>27</sup>. Finally, it has been shown that  $G\alpha_s$ Ras and  $G\alpha_s$ AH can form a functional GTPase when expressed together as separate proteins<sup>28</sup>. Therefore, it is perhaps not surprising that  $G\alpha_s$ AH is displaced relative to  $G\alpha_s$ Ras; however, its location in this crystal structure most probably reflects only one of an ensemble of conformations that it can adopt under physiological conditions, but has been stabilized by crystal packing interactions (Supplementary Fig. 5).

A potential concern is that Nb35, which was used to facilitate crystallization, may be responsible for the displacement of the  $G\alpha_s$ AH. However, Nb35, which binds at the interface between  $G\beta$  and  $G\alpha_s$ Ras (Fig. 2b and Supplementary Fig. 6), would not be expected to interact

with the  $G\alpha_s$ AH or interfere with its interactions with  $G\alpha_s$ Ras. None of the Nb35 contacts on the Ras domain are involved in interactions with  $G\alpha_s$ AH on the basis of the crystal structure of  $G\alpha_s$ –GTP $\gamma$ S (1AZT). Moreover, if we superimpose the structures of the Ras domains of  $G\alpha_s$ –GTP $\gamma$ S (1AZT) and  $\beta_2$ AR–Gs, there is no overlap between Nb35 and the  $\alpha$ -helical domain of  $G\alpha_s$ –GTP $\gamma$ S (Supplementary Fig. 6). Similarly, if we align the  $G\beta$  subunits of the Gi–GDP heterotrimer (1GP2) with that of  $\beta_2$ AR–Gs, there is no overlap between Nb35 and the  $\alpha$ -helical domain of Gi (Supplementary Fig. 6). This analysis is in agreement with single particle electron microscopy studies which provide further evidence that Nb35 does not disrupt interactions between  $G\alpha_s$ AH and  $G\alpha_s$ Ras (Westfield *et al.*, manuscript submitted).

The conformational links between the  $\beta_2$ AR and the nucleotide-binding pocket primarily involve the amino and carboxy-terminal helices of  $G\alpha_s$  (Fig. 4). Figure 5b focuses on the region of  $G\alpha_s$ Ras that undergoes the largest conformational change when comparing the structure of  $G\alpha_s$ Ras from the  $\beta_2$ AR–Gs complex with that from the  $G\alpha_s$ –GTP $\gamma$ S complex<sup>26</sup>. The largest difference is observed for the  $\alpha$ 5-helix, which is displaced 6 Å towards the receptor and rotated as the carboxy-terminal end projects into transmembrane core of the  $\beta_2$ AR. Previous studies using a variety of approaches have demonstrated the important role of the  $\alpha$ 5-helix in GPCR–G protein interactions<sup>29,30</sup>. Associated with movement of the  $\alpha$ 5-helix, the  $\beta$ 6– $\alpha$ 5 loop, which interacts with the guanine ring in the  $G\alpha_s$ –GTP $\gamma$ S structure, is displaced outward, away from the nucleotide-binding pocket (Fig. 5b–d). The movement of  $\alpha$ 5-helix is also associated with changes in interactions between this helix and the  $\beta$ 6-strand, the  $\alpha$ N– $\beta$ 1 loop, and the  $\alpha$ 1-helix. The  $\beta$ 1-strand forms another link between the  $\beta_2$ AR and the nucleotide-binding pocket. The carboxy-terminal end of this strand changes conformation around Gly 47, and there are further changes in the  $\beta$ 1– $\alpha$ 1 loop (P-loop) that coordinates the  $\beta$ -phosphate in the GTP-bound form (Fig. 5b–d). The observations in the crystal structure are in agreement with deuterium exchange experiments where there is enhanced deuterium exchange in the  $\beta$ 1-strand and the amino-terminal end of the  $\alpha$ 5-helix upon formation of the nucleotide-free  $\beta_2$ AR–Gs complex (Chung *et al.*, manuscript submitted). The deuterium exchange-mass spectrometry (DXMS) studies provide additional insights into the dynamic nature of these conformational changes in Gs upon complex formation (Chung *et al.*, manuscript submitted).

The structure of a GDP-bound Gs heterotrimer has not been determined, so it is not possible to directly compare the  $G\alpha_s$ – $G\beta\gamma$  interface



**Figure 5 | Conformational changes in  $G\alpha_s$ .** a, A comparison of  $G\alpha_s$  in the  $\beta_2$ AR–Gs complex (orange) with the GTP $\gamma$ S-bound  $G\alpha_s$  (grey)<sup>26</sup> (PDB ID: 1AZT). GTP $\gamma$ S is shown as spheres. The helical domain of  $G\alpha_s$  ( $G\alpha_s$ AH) shows a marked displacement relative to its position in the GTP $\gamma$ S-bound state. b, The  $\alpha$ 5-helix of  $G\alpha_s$  is rotated and displaced towards the  $\beta_2$ AR, perturbing the

$\beta$ 6– $\alpha$ 5 loop which otherwise forms part of the GTP $\gamma$ S-binding pocket. c, The  $\beta$ 1– $\alpha$ 1 loop (P-loop) and  $\beta$ 6– $\alpha$ 5 loop of  $G\alpha_s$  interact with the phosphates and purine ring, respectively, of GTP $\gamma$ S in the  $G\alpha_s$ –GTP $\gamma$ S structure. d, The  $\beta$ 1– $\alpha$ 1 and  $\beta$ 6– $\alpha$ 5 loops are rearranged in the nucleotide-free  $\beta_2$ AR–Gs structure.

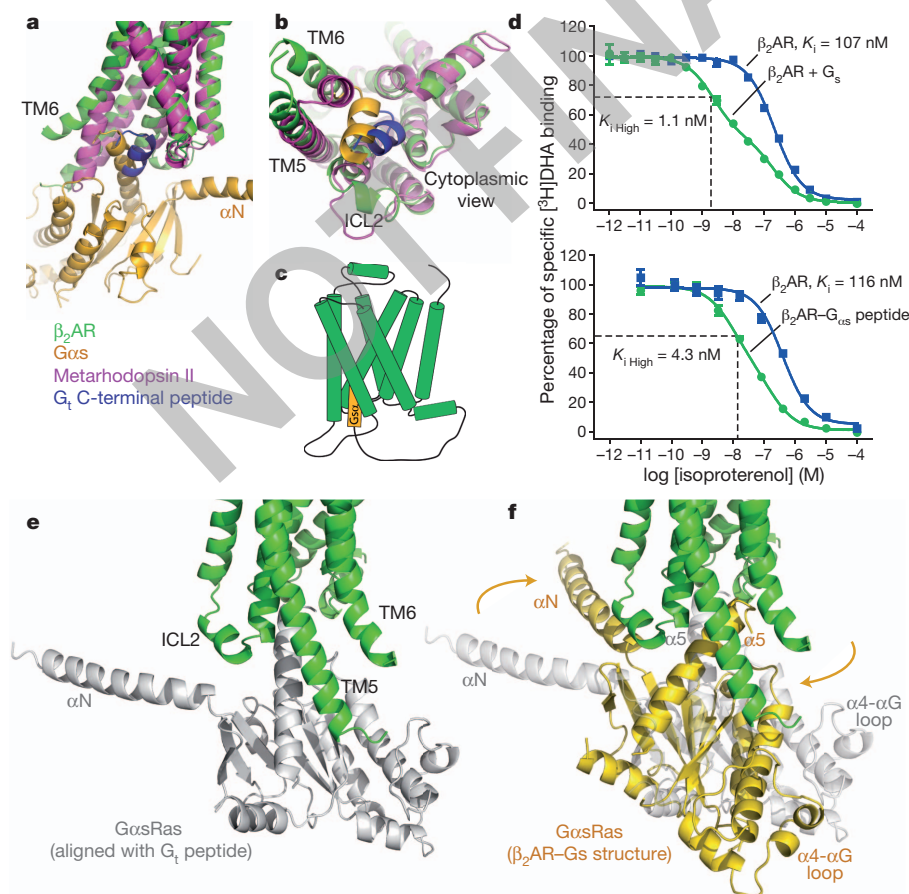
before and after formation of the  $\beta_2$ AR–Gs complex. On the basis of the structure of the GDP-bound Gi heterotrimer<sup>31</sup>, we do not observe large changes in interactions between G $\alpha$ sRas and G $\beta\gamma$  upon formation of the complex with  $\beta_2$ AR. This is also consistent with deuterium exchange studies (Chung *et al.*, manuscript submitted). As discussed above, Nb35 binds at the interface between G $\alpha$ sRas and G $\beta$  (Fig. 2b); therefore, we cannot exclude the possibility that Nb35 may influence the relative orientation of the G $\alpha$ sRas–G $\beta\gamma$  interface in the crystal structure.

### Assembly of the $\beta_2$ AR–Gs complex

Clues to the initial stages of complex formation may come from the recent active state structures of rhodopsin<sup>32,33</sup>. Figure 6a, b compares the active-state structure of  $\beta_2$ AR in the  $\beta_2$ AR–Gs complex with the recent structure of metarhodopsin II bound to a peptide representing the carboxy terminus of transducin<sup>32</sup>. The conformational changes in TM5 and TM6 are smaller in metarhodopsin II, and the position of the carboxy-terminal  $\alpha$  helix of transducin is tilted by approximately 30° relative to the position of the homologous region of Gs. These may represent fundamental differences in the receptor–G protein interactions between these two proteins, but given the conservation of the G-protein binding pocket, the changes more probably reflect the more extensive contacts formed with the intact G protein. The position of the transducin peptide in metarhodopsin II may represent the initial interaction between a GDP-bound G protein and a GPCR. We have attempted to reproduce a similar complex between the  $\beta_2$ AR and a synthetic peptide representing the carboxy-terminal 20 amino acids of Gs, but did not observe any effect of this peptide on receptor function, possibly due to the solubility and behaviour of the peptide in solution. However, when the carboxy-terminal 20 amino acids of Gs are fused to the carboxy terminus of the  $\beta_2$ AR (Fig. 6c), we observe a 27-fold increase in agonist affinity (Fig. 6d). This effect is only 3.5-fold smaller than the effect we observe on agonist binding affinity

in the  $\beta_2$ AR–Gs complex, and demonstrates that there is a functional interaction between the peptide and receptor that may represent an initial stage in  $\beta_2$ AR–Gs complex formation. Figure 6e and f presents a possible sequence of interactions of  $\beta_2$ AR and Gs when forming the nucleotide-free complex. The first interaction of the  $\beta_2$ AR with the Gs heterotrimer would require a movement of the carboxy terminus of the  $\alpha$ 5-helix away from the  $\beta$ 6-strand to permit interactions with the  $\beta_2$ AR similar to those observed in metarhodopsin II (Fig. 6e). The availability of the carboxy terminus of the  $\alpha$ 5-helix for interactions with the  $\beta_2$ AR is supported by deuterium exchange studies (Chung *et al.*, manuscript submitted) showing that this segment is more dynamic in the Gs–GDP heterotrimer than would be expected from the crystal structure of G $\alpha$ s<sup>26</sup>. The subsequent formation of more extensive interactions between the  $\beta_2$ AR ICL2 and the amino terminus of G $\alpha$ s requires a rotation of G $\alpha$ sRas relative to the receptor and would be associated with further conformational changes in both  $\beta_2$ AR and G $\alpha$ sRas (Fig. 6f). We cannot say when GDP is released during the formation of the complex; however, we speculate that uncoupling of the G $\alpha$ sAH from G $\alpha$ sRas is a consequence of nucleotide release or at least a coincident event. This binding model is in agreement with deuterium exchange experiments (Chung *et al.*, manuscript submitted).

The  $\beta_2$ AR–Gs complex crystal structure provides the first high-resolution view of transmembrane signalling for a GPCR. We now have a framework to design experiments to investigate the mechanism of complex formation, GTP binding and complex dissociation. Of particular interest will be studies designed to determine the functional significance of the large movement of G $\alpha$ sAH relative to G $\alpha$ sRas that is observed in the  $\beta_2$ AR–Gs complex. A better understanding of the structural basis for G protein activation may provide new approaches for drug discovery. The high degree of structural homology within the ligand-binding pocket has posed challenges for developing highly selective drugs for specific GPCR targets. In contrast, there is relatively



**Figure 6 | Possible sequence of  $\beta_2$ AR–Gs complex formation.** **a**, **b**, Comparison of the  $\beta_2$ AR–Gs structure (green and orange) with metarhodopsin II<sup>32</sup> (PDB ID: 3PQR) (purple) bound with the carboxy-terminal peptide of transducin (G $\gamma$ ) (blue). TM7 has been omitted in panel a to better visualize the G proteins. **c**, Cartoon of the  $\beta_2$ AR–G $\alpha$ s peptide fusion construct used in the binding experiments. **d**, Competition binding experiments between [<sup>3</sup>H]dihydroalprenolol ([<sup>3</sup>H]DHA) and full agonist isoproterenol. Top panel shows binding data (reproduced from ref. 12) on  $\beta_2$ AR reconstituted in HDL particles with and without Gs heterotrimer. The fraction of  $\beta_2$ AR in the  $K_{i\text{ High}}$  state for the  $\beta_2$ AR with Gs is 0.55. Bottom panel shows binding to  $\beta_2$ AR and a  $\beta_2$ AR–G $\alpha$ s peptide fusion expressed in Sf9 cell membranes. The fraction of  $\beta_2$ AR in the  $K_{i\text{ High}}$  state for the  $\beta_2$ AR–G $\alpha$ s peptide fusion is 0.68. **e**, The initial interaction of agonist-bound  $\beta_2$ AR and G $\alpha$ sRas may involve an orientation of the carboxy terminus of G $\alpha$ sRas similar to that of the carboxy-terminal peptide of transducin in the structure of metarhodopsin II. **f**, The final position of G $\alpha$ sRas on the  $\beta_2$ AR as observed in the  $\beta_2$ AR–Gs complex.



low homology at the interface between the  $\beta_2$ AR and G $\alpha$ s, so identifying sequence and structural features that define specificity for particular G proteins may enable the development of selective inhibitors of specific GPCR–G protein interactions.

## METHODS SUMMARY

The  $\beta_2$ AR–Gs complex was crystallized from  $\beta_2$ AR and Gs protein expressed in insect cells. Crystallography was aided by fusing T4 lysozyme to the amino terminus of the  $\beta_2$ AR and the addition of a nanobody (Nb35) that binds at the interface between the G $\alpha$  and G $\beta$  subunits. Crystals were grown in a lipidic cubic phase using MAG7.7, a lipid that accommodates membrane proteins with larger hydrophilic surfaces<sup>15</sup>. Diffraction data were measured at beamline 23ID-B of the Advanced Photon Source and the structure was solved by molecular replacement. For more experimental details see Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 8 June; accepted 11 July 2011.**

**Published online 19 July 2011.**

- Dixon, R. A. *et al.* Cloning of the gene and cDNA for mammalian  $\beta$ -adrenergic receptor and homology with rhodopsin. *Nature* **321**, 75–79 (1986).
- Rasmussen, S. G. *et al.* Crystal structure of the human  $\beta_2$  adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into  $\beta_2$ -adrenergic receptor function. *Science* **318**, 1266–1273 (2007).
- Lefkowitz, R. J. Seven transmembrane receptors: something old, something new. *Acta Physiol. (Oxf.)* **190**, 9–19 (2007).
- Brandt, D. R., Asano, T., Pedersen, S. E. & Ross, E. M. Reconstitution of catecholamine-stimulated guanosinetriphosphatase activity. *Biochemistry* **22**, 4357–4362 (1983).
- Cerione, R. A. *et al.* The mammalian  $\beta_2$ -adrenergic receptor: reconstitution of functional interactions between pure receptor and pure stimulatory nucleotide binding protein of the adenylate cyclase system. *Biochemistry* **23**, 4519–4525 (1984).
- Ross, E. M., Maguire, M. E., Sturgill, T. W., Biltonen, R. L. & Gilman, A. G. Relationship between the  $\beta$ -adrenergic receptor and adenylate cyclase. Studies of ligand binding and enzyme activity in purified membranes of S49 lymphoma cells. *J. Biol. Chem.* **252**, 5761–5775 (1977).
- De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled  $\beta$ -adrenergic receptor. *J. Biol. Chem.* **255**, 7108–7117 (1980).
- Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Azzi, M. *et al.*  $\beta$ -arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **100**, 11406–11411 (2003).
- Lefkowitz, R. J. & Shenoy, S. K. Transduction of receptor signals by  $\beta$ -arrestins. *Science* **308**, 512–517 (2005).
- Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the  $\beta_2$  adrenoceptor. *Nature* **469**, 175–180 (2011).
- Chae, P. S. *et al.* Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature Methods* **7**, 1003–1008 (2010).
- Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66**, 639–678 (1997).
- Misquitta, L. V. *et al.* Membrane protein crystallization in lipidic mesophases with tailored bilayers. *Structure* **12**, 2113–2124 (2004).
- Domanska, K. *et al.* Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic  $\beta_2$ -microglobulin variant. *Proc. Natl Acad. Sci. USA* **108**, 1314–1319 (2011).
- Warne, T. *et al.* Structure of a  $\beta_1$ -adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).
- Hofmann, K. P. *et al.* A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34**, 540–552 (2009).
- Moro, O., Lameh, J., Hogger, P. & Sadée, W. Hydrophobic amino acid in the i2 loop plays a key role in receptor–G protein coupling. *J. Biol. Chem.* **268**, 22273–22276 (1993).
- Baltensperger, K. *et al.* The  $\beta$ -adrenergic receptor is a substrate for the insulin receptor tyrosine kinase. *J. Biol. Chem.* **271**, 1061–1064 (1996).
- Jastrzebska, B., Tsybovsky, Y. & Palczewski, K. Complexes between photoactivated rhodopsin and transducin: progress and questions. *Biochem. J.* **428**, 1–10 (2010).
- Johnston, C. A. & Siderovski, D. P. Receptor-mediated activation of heterotrimeric G-proteins: current structural insights. *Mol. Pharmacol.* **72**, 219–230 (2007).
- Breitwieser, G. E. G protein-coupled receptor oligomerization: implications for G protein activation and cell signaling. *Circ. Res.* **94**, 17–27 (2004).
- Banères, J. L. & Parello, J. Structure-based analysis of GPCR function: evidence for a novel pentameric assembly between the dimeric leukotriene B4 receptor BLT1 and the G-protein. *J. Mol. Biol.* **329**, 815–829 (2003).
- Angers, S., Salahpour, A. & Bouvier, M. Dimerization: an emerging concept for G protein-coupled receptor ontogeny and function. *Annu. Rev. Pharmacol. Toxicol.* **42**, 409–435 (2002).
- Sunahara, R. K., Tesmer, J. J., Gilman, A. G. & Sprang, S. R. Crystal structure of the adenylyl cyclase activator G $\alpha_s$ . *Science* **278**, 1943–1947 (1997).
- Van Eps, N. *et al.* Interaction of a G protein with an activated receptor opens the interdomain interface in the alpha subunit. *Proc. Natl Acad. Sci. USA* **108**, 9420–9424 (2011).
- Markby, D. W., Onrust, R. & Bourne, H. R. Separate GTP binding and GTPase activating domains of a G alpha subunit. *Science* **262**, 1895–1901 (1993).
- Conklin, B. R. & Bourne, H. R. Structural elements of G $\alpha$  subunits that interact with G $\beta\gamma$ , receptors, and effectors. *Cell* **73**, 631–641 (1993).
- Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Rev. Mol. Cell Biol.* **9**, 60–71 (2008).
- Wall, M. A. *et al.* The structure of the G protein heterotrimer G $\alpha_{i1}\beta_1\gamma_2$ . *Cell* **83**, 1047–1058 (1995).
- Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651–655 (2011).
- Standfuss, J. *et al.* The structural basis of agonist-induced activation in constitutively active rhodopsin. *Nature* **471**, 656–660 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge support from National Institutes of Health Grants NS028471 (B.K.K.) and GM083118 (B.K.K. and R.K.S.), GM56169 (W.I.W.), P01 GM75913 (S.H.G.) and P60DK-20572 (R.K.S.), GM75915, P50GM073210 and U54GM094599 (M.C.), Science Foundation Ireland (07/IN.1/B1836) and FP7 COST Action CM0902 (M.C.), the Mathers Foundation (B.K.K. and W.I.W.), the Lundbeck Foundation (Junior Group Leader Fellowship, S.G.F.R.), the University of Michigan Biomedical Sciences Scholars Program (R.K.S.), the Fund for Scientific Research of Flanders (FWO-Vlaanderen) and the Institute for the encouragement of Scientific Research and Innovation of Brussels (ISIRB) (E.P. and J.S.), The Danish Council for Independent Research, Medical Sciences (J.M.M.). We thank A. Coughlan for help with lipid synthesis.

**Author Contributions** S.G.F.R. performed the final stages of  $\beta_2$ AR purification; assisted with  $\beta_2$ AR and Gs protein virus production and expression in insect cell cultures; worked out conditions to form and stabilize the  $\beta_2$ AR–Gs complex following screening, identification and characterization of the BI-167107 agonist and MNG-3 detergent; developed the  $\beta_2$ AR–Gs complex purification strategy with B.K.K. and characterized the stability of the complex under a variety of conditions; purified and analysed all preparations of the  $\beta_2$ AR–Gs complex used for crystallography, DXMS and electron microscopy studies, immunization, and nanobody selection; expressed and purified nanobodies and characterized  $\beta_2$ AR–Gs–Nb binding by size exclusion chromatography; set up crystallization trials in detergent solution, bicelles and lipidic cubic phase; crystallized the T4L– $\beta_2$ AR–Gs, T4L– $\beta_2$ AR–Gs–Nb37 and T4L– $\beta_2$ AR–Gs–Nb35 complexes; optimized crystallization conditions and grew crystals for data collection; assisted with data collection and manuscript preparation. B.T.D. managed Gs heterotrimer subunit virus production and titration; expressed and purified Gs protein; with R.K.S. he identified the use of apyrase in forming the  $\beta_2$ AR–Gs complex and foscarnet/pyrophosphate during crystallography; reconstituted the  $\beta_2$ AR–Gs complex and receptor alone in high density lipoprotein (HDL) particles which were used for the initial nanobody selection. He assisted with data collection. Y.Z. designed, generated and optimized the T4L– $\beta_2$ AR fusion protein, characterized its expression and functional properties, and prepared baculovirus for large scale expression. A.C.K. collected crystals, collected and processed diffraction data, solved and refined the structure, and assisted with manuscript preparation. K.Y.C. developed the cross-linking conditions for the purified  $\beta_2$ AR–Gs complex used for immunization of llamas. E.P. performed llama immunization, cloned and expressed nanobodies and performed initial selections. J.S. supervised nanobody production. D.C. assisted with Gs heterotrimer expression and purification. J.M.M. generated the  $\beta_2$ AR–Gs $\alpha$  peptide fusion construct, expressed it in insect cell membranes and performed competition binding experiments. F.S.T. expressed  $\beta_2$ AR in insect cell cultures and with T.S.K. performed the initial stage of  $\beta_2$ AR purification. S.T.A.S., J.A.L. and M.C. provided the 7.7 MAG lipid and helpful suggestions for lipidic mesophase crystallization using this lipid. P.S.C. and S.H.G. provided MNG-3 detergent for stabilization of the  $\beta_2$ AR–Gs complex. G.S. provided the essential feedback from the electron microscopy studies that helped guide the crystallization effort. W.I.W. oversaw data processing, structure determination and refinement. R.K.S. supervised Gs protein production, provided valuable ideas and insights into Gs structure and function, and assisted with data collection and manuscript preparation. B.K.K. was responsible for overall project strategy and management, harvested crystals and assisted with collection of diffraction data, and wrote the manuscript.

**Author Information** Coordinates and structure factors for the  $\beta_2$ AR–Gs complex are deposited in the Protein Data Bank (accession code 3SN6). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to B.K.K. (kobilka@stanford.edu) or R.K.S. (sunahara@umich.edu).



## METHODS

**Expression and purification of  $\beta_2$ AR, Gs heterotrimer and nanobody-35.** An amino-terminally fused T4 lysozyme- $\beta_2$ AR construct with  $\beta_2$ AR truncated in position 365 (T4L- $\beta_2$ AR, described in detail below) was expressed in *Sf9* insect cell cultures infected with recombinant baculovirus (BestBac, Expression Systems), and solubilized in *n*-dodecyl- $\beta$ -D-maltoside (DDM) according to methods described previously<sup>34</sup> (see Supplementary Fig. 7 for purification overview). A  $\beta_2$ AR construct truncated after residue 365 ( $\beta_2$ AR-365) was used for the majority of the analytical experiments. M1 Flag affinity chromatography (Sigma) served as the initial purification step followed by alprenolol-Sepharose chromatography for selection of functional receptor. A subsequent M1 Flag affinity chromatography step was used to exchange receptor-bound alprenolol for high-affinity agonist BI-167107. The agonist-bound receptor was eluted, dialysed against buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 0.1% DDM and 10  $\mu$ M BI-167107), treated with lambda phosphatase (New England Biolabs), and concentrated to approximately 50 mg ml<sup>-1</sup> with a 50 kDa molecular weight cut off (MWCO) Millipore concentrator. Prior to spin concentration, the  $\beta_2$ AR-365 construct, but not T4L- $\beta_2$ AR, was treated with PNGaseF (New England Biolabs) to remove amino-terminal amino-linked glycosylation. The purified receptor was routinely analysed by SDS-PAGE/Coomassie brilliant blue staining (see Supplementary Fig. 8a).

Bovine G $\alpha_s$  short, His<sub>6</sub>-rat G $\beta_1$  and bovine G $\gamma_2$  were expressed in HighFive insect cells (Invitrogen) grown in Insect Xpress serum-free media (Lonza). Cultures were grown to a density of 1.5 million cells per ml and then infected with three separate *Autographa californica* nuclear polyhedrosis virus each containing the gene for one of the G protein subunits at a 1:1 multiplicity of infection (the viruses were a gift from A. Gilman). After 40–48 h of incubation the infected cells were harvested by centrifugation and resuspended in 75 ml lysis buffer (50 mM HEPES, pH 8.0, 65 mM NaCl, 1.1 mM MgCl<sub>2</sub>, 1 mM EDTA, 1  $\times$  PTT (35  $\mu$ g ml<sup>-1</sup> phenylmethanesulphonyl fluoride, 32  $\mu$ g ml<sup>-1</sup> tosyl phenylalanyl chloromethyl ketone, 32  $\mu$ g ml<sup>-1</sup> tosyl lysyl chloromethyl ketone), 1  $\times$  LS (3.2  $\mu$ g ml<sup>-1</sup> leupeptin and 3.2  $\mu$ g ml<sup>-1</sup> soybean trypsin inhibitor), 5 mM  $\beta$ -mercaptoethanol ( $\beta$ -ME), and 10  $\mu$ M GDP) per litre of culture volume. The suspension was pressurized with 600 p.s.i. N<sub>2</sub> for 40 min in a nitrogen cavitation bomb (Parr Instrument Company). After depressurization, the lysate was centrifuged to remove nuclei and unlysed cells, and then ultracentrifuged at 180,000g for 40 min. The pelleted membranes were resuspended in 30 ml wash buffer (50 mM HEPES, pH 8.0, 50 mM NaCl, 100  $\mu$ M MgCl<sub>2</sub>, 1  $\times$  PTT, 1  $\times$  LS, 5 mM  $\beta$ -ME, 10  $\mu$ M GDP) per litre culture volume using a Dounce homogenizer and centrifuged again at 180,000g for 40 min. The washed pellet was resuspended in a minimal volume of wash buffer and flash-frozen with liquid nitrogen.

The frozen membranes were thawed and diluted to a total protein concentration of 5 mg ml<sup>-1</sup> with fresh wash buffer. Sodium cholate detergent was added to the suspension at a final concentration of 1.0%, MgCl<sub>2</sub> was added to a final concentration of 5 mM, and 0.05 mg of purified protein phosphatase 5 (prepared in house) was added per litre of culture volume. The sample was stirred on ice for 40 min, and then centrifuged at 180,000g for 40 min to remove insoluble debris. The supernatant was diluted fivefold with Ni-NTA load buffer (20 mM HEPES, pH 8.0, 363 mM NaCl, 1.25 mM MgCl<sub>2</sub>, 6.25 mM imidazole, 0.2% Anzergent 3-12, 1  $\times$  PTT, 1  $\times$  LS, 5 mM  $\beta$ -ME, 10  $\mu$ M GDP), taking care to add the buffer slowly to avoid dropping the cholate concentration below its critical micelle concentration too quickly. Ni-NTA resin (3 ml; Qiagen) pre-equilibrated in Ni-NTA wash buffer 1 (20 mM HEPES, pH 8.0, 300 mM NaCl, 2 mM MgCl<sub>2</sub>, 5 mM imidazole, 0.2% cholate, 0.15% Anzergent 3-12, 1  $\times$  PTT, 1  $\times$  LS, 5 mM  $\beta$ -ME, 10  $\mu$ M GDP) per litre culture volume was added and the sample was stirred on ice for 20 min. The resin was collected into a gravity column and washed with 4  $\times$  column volumes of Ni-NTA wash buffer 1, Ni-NTA wash buffer 2 (20 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 10 mM imidazole, 0.15% Anzergent 3-12, 0.1% DDM, 1  $\times$  PTT, 1  $\times$  LS, 5 mM  $\beta$ -ME, 10  $\mu$ M GDP), and Ni-NTA wash buffer 3 (20 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 5 mM imidazole, 0.1% DDM, 1  $\times$  PTT, 1  $\times$  LS, 5 mM  $\beta$ -ME, 10  $\mu$ M GDP). The protein was eluted with Ni-NTA elution buffer (20 mM HEPES, pH 8.0, 40 mM NaCl, 1 mM MgCl<sub>2</sub>, 200 mM imidazole, 0.1% DDM, 1  $\times$  PTT, 1  $\times$  LS, 5 mM  $\beta$ -ME, 10  $\mu$ M GDP). Protein-containing fractions were pooled and MnCl<sub>2</sub> was added to a final concentration of 100  $\mu$ M. Purified lambda protein phosphatase (50  $\mu$ g; prepared in house) was added per litre of culture volume and the eluate was incubated on ice with stirring for 30 min. The eluate was passed through a 0.22- $\mu$ m filter and loaded directly onto a MonoQ HR 16/10 column (GE Healthcare) equilibrated in MonoQ buffer A (20 mM HEPES, pH 8.0, 50 mM NaCl, 100  $\mu$ M MgCl<sub>2</sub>, 0.1% DDM, 5 mM  $\beta$ -ME, 1  $\times$  PTT). The column was washed with 150 ml buffer A at 5 ml min<sup>-1</sup> and bound proteins were eluted over 350 ml with a linear gradient up to 28% MonoQ buffer B (same as buffer A except with 1 M NaCl). Fractions were collected in tubes spotted with enough GDP to make a final concentration of

10  $\mu$ M. The Gs-containing fractions were concentrated to 2 ml using a stirred ultrafiltration cell (Amicon) with a 10-kDa nominal molecular weight limit (NMWL) regenerated cellulose membrane (Millipore). The concentrated sample was run on a Superdex 200 prep grade XK 16/70 column (GE Healthcare) equilibrated in S200 buffer (20 mM HEPES, pH 8.0, 100 mM NaCl, 1.1 mM MgCl<sub>2</sub>, 1 mM EDTA, 0.012% DDM, 100  $\mu$ M TCEP, 2  $\mu$ M GDP). The fractions containing pure Gs were pooled, glycerol was added to 10% final concentration, and then the protein was concentrated to at least 10 mg ml<sup>-1</sup> using a 30 kDa MWCO centrifugal ultrafiltration device (Millipore). The concentrated sample was then aliquoted, flash frozen, and stored at -80 °C. A typical yield of final, purified Gs heterotrimer from 8 l of cell culture volume was 6 mg.

Nanobody-35 (Nb35) was expressed in the periplasm of *E. coli* strain WK6, extracted, and purified by nickel affinity chromatography according to previously described methods<sup>12</sup> followed by ion-exchange chromatography (Supplementary Fig. 9a) using a Mono S 10/100 GL column (GE Healthcare). Selected Nb35 fractions were dialysed against buffer (10 mM HEPES, pH 7.5, 100 mM NaCl) and concentrated to approximately 65 mg ml<sup>-1</sup> with a 10 kDa MWCO Millipore concentrator.

**Complex formation, stabilization and purification.** Formation of a stable complex (see Supplementary Fig. 10) was accomplished by mixing Gs heterotrimer at approximately 100  $\mu$ M concentration with BI-167107-bound T4L- $\beta_2$ AR (or  $\beta_2$ AR-365) in molar excess (approximately 130  $\mu$ M) in 2 ml buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.1% DDM, 1 mM EDTA, 3 mM MgCl<sub>2</sub>, 10  $\mu$ M BI-167107) and incubating for 3 h at room temperature. BI-167107, which was identified from screening and characterizing approximately 50 different  $\beta_2$ AR agonists (data not shown), has a dissociation half-time of approximately 30 h, providing higher degree of stabilization to the active G protein-bound receptor than other full agonists such as isoproterenol<sup>12</sup>. To maintain the high-affinity nucleotide-free state of the complex, apyrase (25 mU ml<sup>-1</sup>, NEB) was added after 90 min to hydrolyse residual GDP released from G $\alpha_s$  upon binding to the receptor. GMP resulting from hydrolysis of GDP by apyrase has very poor affinity for the G protein in the complex. Rebinding of GDP can cause dissociation of the  $\beta_2$ AR-Gs complex (Supplementary Fig. 1a).

The  $\beta_2$ AR-Gs complex in DDM shows significant dissociation after 48 h at 4 °C (Supplementary Fig. 11a). We screened and characterized over 50 amphiphiles (data not shown) and identified MNG-3 (refs 12, 13; NG-310, Affymetrix-Anatrace) and its closely related analogues as detergents that substantially stabilize the complex (Supplementary Fig. 11a). The complex was exchanged into MNG-3 by adding the  $\beta_2$ AR-Gs mixture (2 ml) to 8 ml buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10  $\mu$ M BI-167107) containing 1% MNG-3 for 1 h at room temperature.

At this stage the mixture contains the  $\beta_2$ AR-Gs complex, non-functional Gs, and an excess of  $\beta_2$ AR. To separate functional  $\beta_2$ AR-Gs complex from non-functional Gs, and to complete the detergent exchange, the  $\beta_2$ AR-Gs complex was immobilized on M1 Flag resin and washed in buffer (20 mM HEPES, pH 7.5, 100 mM NaCl, 10  $\mu$ M BI-167107, and 3 mM CaCl<sub>2</sub>) containing 0.2% MNG-3. To prevent cysteine bridge-mediated aggregation of  $\beta_2$ AR-Gs complexes, 100  $\mu$ M TCEP was added to the eluted protein before concentrating it with a 50 kDa MWCO Millipore concentrator. Of note, it was discovered later that crystal growth improved at even higher TCEP concentrations (above 1 mM) compared to 100  $\mu$ M TCEP, and that the integrity of the  $\beta_2$ AR-Gs complex in MNG-3 was stable to 10 mM TCEP as measured by gel filtration analysis (Supplementary Fig. 12). In contrast, DDM-solubilized  $\beta_2$ AR loses its ability to bind the high-affinity antagonist [<sup>3</sup>H]dihydroalprenolol ([<sup>3</sup>H]DHA) in 10 mM TCEP (data not shown), probably due to disruption of extracellular disulphide bonds. Iodoacetamide could not be used to block reactive cysteines on Gs  $\alpha$  and  $\beta$  subunits as it caused dissociation of the  $\beta_2$ AR-Gs complex (Supplementary Fig. 12b). The final size exclusion chromatography procedure to separate excess free receptor from the  $\beta_2$ AR-Gs complex (Supplementary Fig. 8b) was performed on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with buffer containing 0.02% MNG-3, 10 mM HEPES, pH 7.5, 100 mM NaCl, 10  $\mu$ M BI-167107, and 100  $\mu$ M TCEP. Peak fractions were pooled (Supplementary Fig. 8b) and concentrated to approximately 90 mg ml<sup>-1</sup> with a 100 kDa MWCO Viva-spin concentrator and analysed by SDS-PAGE/Coomassie brilliant blue staining (Supplementary Fig. 8a) and gel filtration (Supplementary Fig. 8c). To confirm a pure, homogeneous, and dephosphorylated preparation, the  $\beta_2$ AR-Gs complex was routinely analysed by ion exchange chromatography (Supplementary Fig. 8d).

**Protein engineering.** To increase the probability of obtaining crystals of the  $\beta_2$ AR-Gs complex we set out to increase the polar surface area on the extracellular side of the receptor using two strategies. The first approach, to generate extracellular binding antibodies, was not successful. The second approach was to replace the flexible and presumably unstructured amino terminus with the globular protein T4 lysozyme (T4L) used previously to crystallize and solve the carazolol-bound receptor<sup>3</sup>. The construct used here (T4L- $\beta_2$ AR) contained the

cleavable signal sequence followed by the M1 Flag epitope (DYKDDDDA), the Tobacco etch virus (TEV) protease recognition sequence (ENLYFQG), bacteriophage T4 lysozyme from N2 through Y161 including C54T and C97A mutations, and a two residue alanine linker fused to the human  $\beta_2$ AR sequence D29 through G365. The PNGaseF-inaccessible glycosylation site of the  $\beta_2$ AR at N187 was mutated to Glu. M96 and M98 in the first extracellular loop were each replaced by Thr to increase the otherwise low expression level of T4L- $\beta_2$ AR. The threonine mutations did not affect ligand binding affinity for [ $^3$ H]DHA, but caused a small, approximately twofold decrease in affinity for isoproterenol (data not shown).

The  $\beta_2$ AR-Gs peptide fusion construct used for [ $^3$ H]DHA competition binding with isoproterenol was constructed from the receptor truncated at position 365 and fused to the last 21 amino acids of the G $\alpha$ s subunit (amino acids 374–394, except for C379A). A Gly-Ser is inserted between the receptor and the peptide. Also an extended TEV protease site (SENLYFQGS) was introduced in the  $\beta_2$ AR between G360 and G361.

**Stabilization of Gs with nanobodies.** From negative stain electron microscopy imaging (Westfield *et al.*, manuscript submitted), we observed that the  $\alpha$ -helical domain of G $\alpha$ s was flexible and therefore possibly responsible for poor crystal quality. Targeted stabilization of this domain was addressed by immunizing two llamas (*Lama glama*) with the bis(sulphosuccinimidyl)glutarate (BS2G, Pierce) cross-linked  $\beta_2$ AR-Gs-BI-167107 ternary complex. Peripheral blood lymphocytes were isolated from the immunized animals to extract total RNA, prepare cDNA and construct a Nanobody phage display library according to published methods<sup>16</sup>. Nb35 and Nb37 were enriched by two rounds of biopanning on the  $\beta_2$ AR-Gs-BI-167107 ternary complex embedded in biotinylated high-density lipoprotein particles<sup>35</sup>. Nb35 and Nb37 were selected for further characterization because they bind the  $\beta_2$ AR-Gs-BI-167107 ternary complex but not the free receptor in an ELISA assay. Nanobody binding to the  $\beta_2$ AR-Gs complex was confirmed by size exclusion chromatography (Supplementary Fig. 1d), and it was noted that both nanobodies protected the complex from dissociation by GTP $\gamma$ S, suggestive of a stabilizing Gs-Nb interaction (Supplementary Fig. 1d).

**Crystallization.** BI-167107 bound T4L- $\beta_2$ AR-Gs complex and Nb35 were mixed in 1:1.2 molar ratio. The small molar excess of Nb35 was verified by analytical gel filtration (see Supplementary Fig. 9b). The mixture incubated for 1 h at room temperature before mixing with 7.7 MAG (provided by M. Caffrey) containing 10% cholesterol (C8667, Sigma) in 1:1 protein solution to lipid ratio (w/w) using the twin-syringe mixing method reported previously<sup>36</sup>. The concentration of T4L- $\beta_2$ AR-Gs-Nb35 complex in 7.7 MAG was approximately 25 mg ml<sup>-1</sup>. We believe the detergent MNG-3 stabilizes the T4L- $\beta_2$ AR-Gs complex during its incorporation into the lipid cubic phase. This may be due to the high affinity of MNG-3 for the receptor. The  $\beta_2$ AR in MNG-3 maintains its structural integrity even when diluted below the CMC of the detergent, in contrast to  $\beta_2$ AR in DDM, which rapidly loses binding activity (Supplementary Fig. 11b). Moreover, MNG-3 improved crystal size and quality, as previously reported<sup>12,13,37</sup>. The protein:lipid mixture was delivered through an LCP dispensing robot (Gryphon, Art Robbins Instruments) in 40 nl drops to either 24-well or 96-well glass sandwich plates and overlaid en-bloc with 0.8  $\mu$ l precipitant solution. Multiple crystallization leads were initially identified using in-house screens partly based on reagents from the StockOptions Salt kit (Hampton Research). Crystals for data collection were grown in 18 to 22% PEG 400, 100 mM MES, pH 6.5 (Supplementary Fig. 1c), 350 to 450 mM potassium nitrate, 10 mM foscarnet (Supplementary Fig. 1b), 1 mM TCEP (Supplementary Fig. 12c), and 10  $\mu$ M BI-167107. Crystals reached full size within 3–4 days at 20 °C and were picked from a sponge-like mesophase and flash-frozen in liquid nitrogen without additional cryoprotectant.

**Microcrystallography data collection and processing.** Diffraction data were measured at the Advanced Photon Source beamline 23 ID-B. Hundreds of crystals were screened, and a final data set was compiled using diffraction wedges of typically 10 degrees from 20 strongly diffracting crystals (Supplementary Table 2). All data reduction was performed using HKL2000 (ref. 38). Although in many cases diffraction to beyond 3 Å was seen in initial frames, radiation damage and anisotropic diffraction resulted in low completeness in higher resolution shells. Analysis of the final data set by the UCLA diffraction anisotropy server<sup>39</sup> indicated that diffraction along the  $a^*$  axis was superior to that in other directions. On

the basis of an  $F/\sigma(F)$  cutoff of 3 along each reciprocal space axis, reflections were subjected to an anisotropic truncation with resolution limits of 2.9, 3.2 and 3.2 Å along  $a^*$ ,  $b^*$  and  $c^*$  before use in refinement. We report this structure to an overall resolution of 3.2 Å. Despite the low completeness in the highest resolution shells (Supplementary Table 3) inclusion of these reflections gave substantial improvements in map quality and lower  $R_{\text{free}}$  during refinement.

**Structure solution and refinement.** The structure was solved by molecular replacement using Phaser<sup>40,41</sup>. The order of the molecular replacement search was found to be critical in solving the structure. In order, the search models used were: the  $\beta$  and  $\gamma$  subunits from a Gi heterotrimer (PDB ID: 1GP2), the Gs  $\alpha$  Ras-like domain (PDB ID: 1AZT), the active-state  $\beta_2$ AR (PDB ID: 3P0G), a  $\beta_2$ AR-binding nanobody (PDB ID: 3P0G), T4 lysozyme (PDB ID: 2RH1), and the Gs  $\alpha$ -helical domain (PDB ID: 1AZT). Following the determination of the initial structure by molecular replacement, rigid body refinement and simulated annealing were performed in Phenix<sup>42</sup> and BUSTER<sup>43</sup>, followed by restrained refinement and manual rebuilding in Coot<sup>44</sup>. After iterative refinement and manual adjustments, the structure was refined in CNS using the DEN method<sup>45</sup>. Although the resolution of this structure exceeds that for which DEN is typically most useful, the presence of several poorly resolved regions indicated that the incorporation of additional information to guide refinement could provide better results. The DEN reference models used were those used for molecular replacement, with the exception of NB35, which was well ordered and for which no higher resolution structure is available. Side chains were omitted from 53 residues for which there was no electron density past CB below a low contour level of 0.7 $\sigma$  in a 2Fo – Fc map. Figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.). MolProbity was used to determine Ramachandran statistics<sup>46</sup>.

**Competition binding.** Membranes expressing the  $\beta_2$ AR or the  $\beta_2$ AR-Gs peptide fusion were prepared from baculovirus-infected Sf9 cells and [ $^3$ H]DHA-binding performed as previously described<sup>47</sup>. For competition binding, membranes were incubated with [ $^3$ H]DHA (1.1 nM final) and increasing concentrations of (–)-isoproterenol for 1 h before harvesting onto GF/B filters. Competition data were fitted to a two-site binding model and isoproterenol high and low  $K_i$ s and fractions calculated using GraphPad prism.

34. Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* **231**, 269–271 (1995).
35. Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
36. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
37. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist- $\beta_2$  adrenoceptor complex. *Nature* **469**, 236–240 (2011).
38. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
39. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
40. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
41. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
42. Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr. D* **61**, 850–855 (2005).
43. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
45. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
46. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
47. Swaminath, G., Steenhuis, J., Kobilka, B. & Lee, T. W. Allosteric modulation of  $\beta_2$ -adrenergic receptor by Zn<sup>2+</sup>. *Mol. Pharmacol.* **61**, 65–72 (2002).